Research on the Influencing Factors of Fan Reviews and Sentiment of We-media in Social Commerce

Jihua Cao[1,2], Jie Li[1]*, Yunfeng Wang[1], Miao Yin[1]

[1]School of Economics and Management, Hebei University of Technology, Tianjin, China

[2]Beihai Campus, Guilin University of Electronic Technology, Beihai, Guangxi, China

*Corresponding author

Abstract:

Fan reviews of We-media in social commerce reflect the fans' sentiment. Researching the number of fan reviews and the influential factors of fan sentiments can help to understand the fan sentiment characteristics and attract their attention. We use crawler software to crawl Li Ziqi's fan reviews on Today's Headlines, and use SnowNLP sentiment analysis to quantify the sentiment attributes of the reviews into sentiment indexes. We construct three types of features: interactive behavior, sentiment attributes and video attributes. We use regression method to analyze the relationship between the number of reviews and sentimental attributes and other characteristics. The results show that the number of fan reviews is directly proportional to the number of broadcasts, the number of likes, and time span; the average sentiment index is directly proportional to the time span and positive reviews, and inversely proportional to the video duration. The results provide a basis for social media and bloggers to manage fan reviews effectively.

Keywords: Social commerce; Online reviews; Sentiment analysis; Reviews management

I. INTRODUCTION

With the development of social networks and We-media, merchants or We-media bloggers interact with consumers and fans online in social networks, promoting the formation of a new commerce model, namely the social commerce model[1]. Social commerce integrates the strengths of traditional E-commerce and social networks to facilitate the trading of goods through means such as social interaction with fans and content generation. 2021 The government work report states that it will accelerate digital development, build digital economic advantages, collaborate to promote digital industrialisation and digital transformation of industries, accelerate the pace of building a digital society and build a digital China. According to the 47th Statistical Report on the Development Status of the Internet in China by the China Internet Network Information Center, as of December 2020, the scale of China's online shopping users reached 782 million and the scale of online payment users reached 854 million[2]. According to the Internet Society of China's "2021 China Social E-Commerce Industry Development Report", the scale of social commerce in China will be 3.7 trillion yuan in 2020[3]. This shows that the development of social commerce is facing unprecedented opportunities and will be in full force in its continued development in

the future.

In the field of social commerce, fan reviews, as a representative type of user opinion-based data[4], are generated directly by fans. Fan reviews express fans' positive or negative attitudes towards various aspects of a product's functionality or performance, reflecting their attention and satisfaction with various aspects of the product's features[5], and provide new opportunities for product designers to understand customer needs[6-7]. Therefore, sentiment analysis of fan reviews using text mining methods can effectively tap into fans' emotional attitudes towards bloggers and their products[8]. In social commerce, the bandwagon effect of online celebrity bloggers is becoming more and more obvious with the popularity of communication technology. According to statistics from the big data trading platform Data Treasure, food blogger Li Ziqi's net profit reached 168 million in 2019, with its profitability exceeding that of 60% of domestic A-share companies. The number of followers it has on major social media has reached nearly 80 million. As a representative of the net celebrity economy in social commerce, Li Ziqi can reflect the operation of social commerce by taking it as the object of study.

This paper takes the social commerce as the research background, selects food blogger's Li Ziqi as the research object, uses SnowNLP sentiment analysis method to quantify the sentiment attributes of his fans' reviews as sentiment index, and constructs three categories of features: interactive behaviour, sentiment attributes and video attributes. Using logistic regression, the relationship between the number of reviews and sentiment attributes and the constructed features was analysed to explore the changes in fan reviews and sentiment attributes in the social commerce environment, and then provide rationalisation suggestions for social media and bloggers.

II. LITERATURE REVIEW AND STUDY DESIGN

2.1 Literature Review

After a systematic review, existing research on social commerce can be divided into four perspectives, including user behaviour, management level, technical support and business information services[9]. Among them, a large number of scholars have done relevant research on user behaviour in social commerce, which is the mainstream of social commerce research. Based on the S-O-R model, Zhou and Chen[10] analysed the influencing factors of users' usage and sharing behaviour through empirical research; Sun et al[11] conducted an empirical study based on the data of WeChat group members and pointed out that the social atmosphere of user groups influenced their purchasing behaviour; Zhou[12] used the in-depth interview method and sorted out the influencing factors of user stickiness based on content analysis; Gan et al[13] Using the structural equation model of PLS, analysed the factors influencing consumer trust from a consumer perspective; Zhang and Wei[14] constructed a model of the formation mechanism of online impulse purchase and used structural equation modeling to conduct empirical analysis, proposing that users' interpersonal interaction through social networks can influence users' desire to purchase; Shao and Hu[15] constructed a model of product adoption based on continuous view and discrete behavior ideas, and used computational experimental methods to They used computational

experimental methods to analyse the influence of factors such as social advertising on product adoption; Ding and Lin[16] used the least squares method to analyse the factors influencing consumers' willingness to continuously share information through questionnaires; Xiang and Wang[17] combined the S-O-R model with motivation theory to empirically analyse the relationship between user intention and user behaviour; Gan and Xu[18] analysed the influence of perceived value on social commerce users' behavioural intentions based on perceived value theory.

In conclusion, most of the studies in social commerce user behaviour have collected user data through questionnaires, while relatively few studies have used fan reviews; most studies have used structural equation models, and fewer have used data analysis methods such as machine learning; most studies have only stayed in the analysis of user purchase behaviour, and fewer have explored the deeper user sentiments. Therefore, in this paper, from the perspective of user behaviour and sentiment in social commerce[19], we obtain the fan reviews of social commerce bloggers through crawler software, use SnowNLP sentiment analysis method to quantify the sentiment attributes of text reviews as sentiment indices, and construct relevant features. The relationship between the number of reviews and sentiment attributes and the constructed features was analysed using logistic regression methods to explore the changes in fan reviews and sentiment attributes in the social commerce environment.

2.2 Research Design

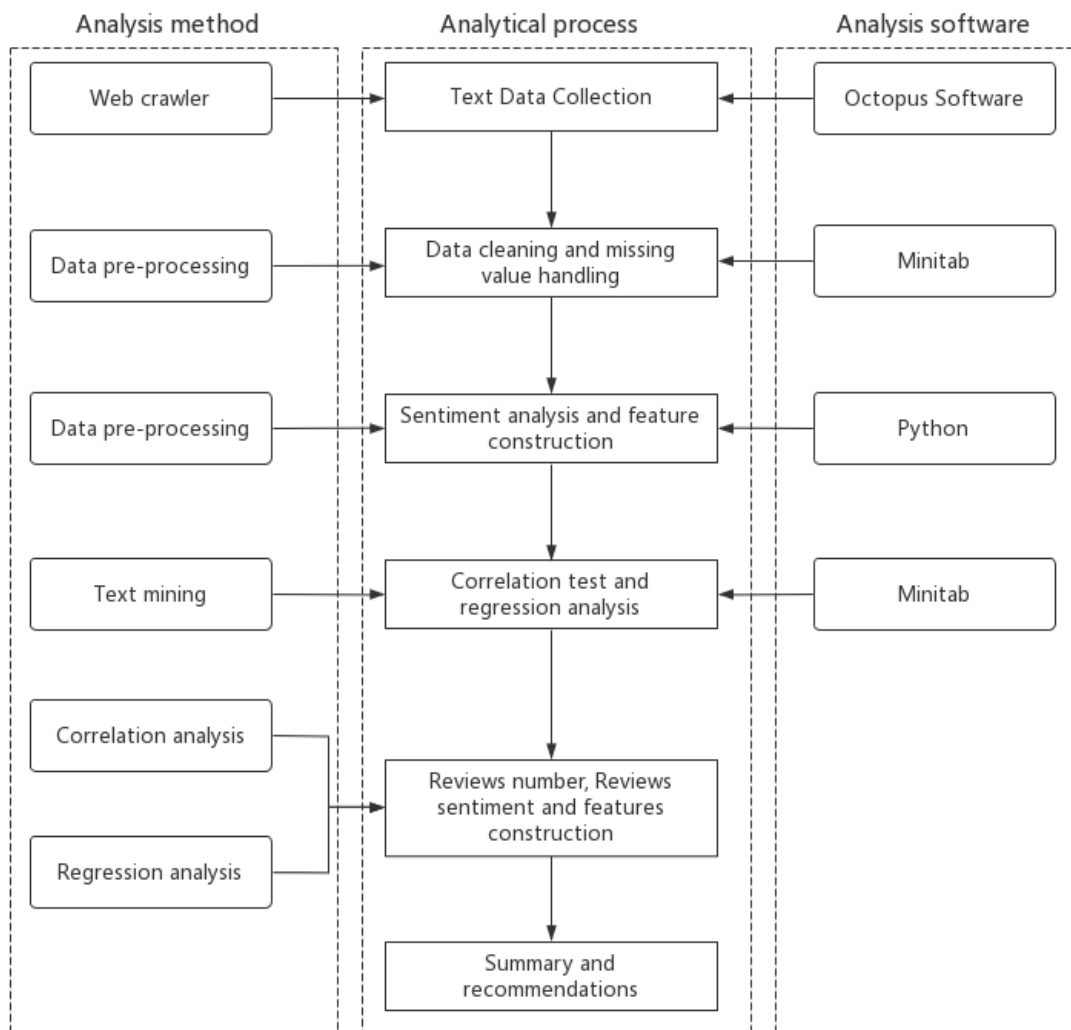The technical roadmap of the study is shown in Fig 1.

Fig 1: Technical roadmap

Firstly, we used the web crawler software "Octopus" to collect the fan reviews from the We-media "Li Ziqi" on social commerce platform Today's Headlines; Secondly, we used the statistical software Minitab to clean the data and process the missing values, and used SnowNLP sentiment analysis to analyse the collected text reviews. Thirdly, we used Minitab to regress the sentiment index of text reviews and the number of fan reviews, with the constructed correlation features, to obtain the regression equation and the results of the study; Finally, we provided recommendations for social media and bloggers on the management of fan reviews.

III. SENTIMENT ANALYSIS AND FEATURE CONSTRUCTION

3.1 Data Collection and Data Pre-processing

The fan reviews of "Li Ziqi", a food blogger on the social commerce platform Today's Headlines, were

selected as the research object, and the crawler software "Octopus" was used to crawl the 78 videos posted by the research object from November 18, 2017 to December 2, 2019. The 139,953 fan reviews were crawled. In order to ensure the quality of the study, videos with less than 500 reviews were excluded. After screening, a total of 122 440 fan reviews from 42 videos were retained for this study. The data information included nine attributes, including video title, video plays, video pose time, fan name, review time, review likes, review replies, video duration and review content.

To ensure the quality of the study, data cleaning and data pre-processing were carried out on the collected review data. Missing values in text reviews and non-text review content such as emoji reviews were screened out. After cleaning and merging the collected data, a total of 120,958 fan text reviews were obtained for use in the study.

3.2 Sentiment Analysis

Sentiment analysis of the crawled text reviews was carried out using SnowNLP, a class library written in Python, a tool library inspired by TexBlob and written specifically for Chinese text content, with text mining functions such as Chinese word separation, lexical annotation, sentiment analysis and text classification, which allows fans to easily perform natural language processing.

With the SnowNLP tool library in the Python language, sentiment analysis was performed on each review crawled, and finally a sentiment index was output after each review. The sentiment index is in the range of 0-1, i.e. when the sentiment index is between 0.5 and 1, the sentiment of the text review is positive and the review is considered positive. Conversely, when the sentiment index of the output is 0-0.5, it means that the sentiment of the review is negative and is recorded as a negative review. When the sentiment index of the output is 0.5, it means that its sentiment intensity is neutral and is recorded as a neutral review.

3.3 Feature Construction

In order to improve the quality of the study, new features were constructed by deeply analysing the relationship between the number of fan reviews and review sentiment with interaction behaviour and video attributes. In addition to the collected data features and the sentiment index of text reviews, 10 features were constructed in this paper for the experiments, and all features were divided into three categories: interactive behaviour, sentiment attributes and video attributes.

The interactive behaviour features include five features: number of plays (10,000 times), number of likes (times), number of reviews (10,000 items), number of replies (10,000 items) and response rate. The number of plays is obtained by crawling and represents the total number of plays of the video; the number of likes of all reviews under each video motion is summed and counted, and is classified as the number of likes feature; the number of reviews under each video motion is summed and counted for each video crawled, and is classified as the number of reviews feature; the number of replies of all reviews under each

video motion is summed and counted, and is classified as the number of replies feature; the number of replies obtained is divided by the number of responses obtained is divided by the number of reviews counted previously to obtain the feature response rate.

The sentiment attribute features include a total of five features: sum of sentiment index, average sentiment index, number of positive reviews (10,000 items), number of negative reviews (10,000 items) and number of neutral reviews (10,000 items). The sentiment index of all reviews under each video dynamic is summed and counted as sentiment index and features; the average sentiment index is obtained by dividing the sum of sentiment index obtained by the number of reviews counted before; the number of reviews with sentiment index greater than 0.5 in each video is counted and summed to obtain the number of positive reviews; similarly, the number of negative reviews and the number of neutral reviews are counted as two features.

The video attribute features include two kinds of features: video duration (minutes) and time span (days). The video duration is obtained by crawling, which represents the number of minutes of video duration; the release time of the last video crawled is set as "origin 0", and then the difference in days between each video and "origin 0" is obtained according to the release time of each video, and the This is classified as the characteristic time span. The results of the above features are shown in TABLE I.

TABLE I. CONSTRUCTED FEATURES AND THEIR MEANINGS

FEATURE CATEGORY	FEATURE	MEANING
INTERACTIVE BEHAVIOUR	NUMBER OF PLAYS	TOTAL NUMBER OF VIDEO PLAYS
	NUMBER OF LIKES	NUMBER OF ALL LIKES
	NUMBER OF REVIEWS	NUMBER OF ALL REVIEWS
	NUMBER OF RESPONSES	NUMBER OF ALL RESPONSES
	RESPONSE RATES	RATIO OF RESPONSES TO REVIEWS
EMOTIONAL ATTRIBUTES	TOTAL SENTIMENT INDEX	SUM OF SENTIMENT INDEXES
	AVERAGE SENTIMENT INDEX	RATIO OF THE SUM OF THE SENTIMENT INDEX TO THE NUMBER OF REVIEWS
	NUMBER OF POSITIVE REVIEWS	NUMBER OF ALL POSITIVE REVIEWS
	NUMBER OF NEGATIVE REVIEWS	NUMBER OF ALL NEGATIVE REVIEWS
	NUMBER OF NEUTRAL REVIEWS	NUMBER OF ALL NEUTRAL REVIEWS
VIDEO ATTRIBUTES	VIDEO DURATION	MINUTES OF VIDEO DURATION
	TIME SPAN	NUMBER OF DAYS THE VIDEO HAS

		BEEN POSTED

IV. RESULT ANALYSIS

The statistical software Minitab was used to regress the number of reviews and the average sentiment index on the other characteristics. The number of reviews was correlated with the above mentioned characteristics, and it was found that the number of reviews was correlated with four characteristics: time span, number of plays, number of likes and response rate. The number of reviews was used as the dependent variable and the remaining four characteristics were used as independent variables for the regression analysis, the results of which are shown in TABLE II. The R^2 of the regression equation was 72.2% and the adjusted R^2 was 69.2%.

TABLE II. REGRESSION ANALYSIS RESULTS OF THE NUMBER OF REVIEWS

INDEPENDENT VARIABLE CHARACTERISTICS	COEFFICIENT	P	T
NUMBER OF PLAYS	5.600	0.019**	2.46
RESPONSE RATES	-1594.900	0.002***	-3.41
NUMBER OF LIKES	0.085	0.000***	4.89
TIME SPAN	4.375	0.000***	4.72

Note: *p<0.1, **<0.05, ***<0.01

The regression equation:

$$\text{Number of reviews} = -666 + 5.600 \times \text{Number of plays} - 1594.900 \times \text{Response rates}$$
$$+ 0.085 \times \text{Number of likes} + 4.38 \times \text{Time span} \tag{1}$$

From the above analysis, it can be seen that the number of reviews is influenced by the number of plays, response rates, number of likes and time span. Firstly, the number of reviews has a significant positive relationship with the number of plays, the number of likes and time span, i.e. the number of reviews increases with the number of plays, the number of likes and time span. This indicates that videos with more plays, more likes and longer releases are more likely to be reviewed on by subsequent fans and attract their attention. Secondly, the number of reviews has a significant negative relationship with the response rate, i.e. the number of reviews decreases as the response rate increases. This suggests that videos with low response rates to all reviews are more likely to be reviewed on by subsequent fans and attract fan attention.

Using the average sentiment index to correlate with the above characteristics, we found that the average sentiment index was correlated with three characteristics: time span, video duration and number of positive reviews. The average sentiment index was used as the dependent variable and the other three characteristics were used as independent variables for the regression analysis, the results of which are shown in TABLE III. The R^2 of the regression equation was 29.4% and the adjusted R^2 was 23.9%.

TABLE III. REGRESSION ANALYSIS RESULTS OF AVERAGE SENTIMENT INDEX

INDEPENDENT VARIABLE CHARACTERISTICS	COEFFICIENT	P	T
TIME SPAN	0.000075	0.035**	2.18
VIDEO DURATION	-0.00563	0.050**	-2.03
NUMBER OF POSITIVE REVIEWS	0.13	0.012**	2.63

Note: *p<0.1, **<0.05, ***<0.01

The regression equation:

$$\text{Average sentiment index} = 0.62 + 0.000075 \times \text{Time span} - 0.00563 \times \text{Video duration}$$
$$+ 0.13 \times \text{Number of positive reviews} \qquad (2)$$

From the above analysis, it can be seen that the average sentiment index is influenced by the time span, the video duration and the number of positive reviews. Firstly, the average sentiment index has a significant positive relationship with the time span and the number of positive reviews, i.e. the average sentiment index increases with the increase in the time span and the number of positive reviews. This indicates that videos with a high number of positive reviews and a long time span have a high tendency to be positively reviewed on by subsequent fans. Secondly, the average sentiment index has a significant negative relationship with video length, i.e. the average sentiment index decreases as the length of the video increases. This suggests that videos with a short video duration have a high positive tendency for subsequent fan reviews.

V. CONCLUSIONS AND RECOMMENDATIONS

In social commerce, the number of fan reviews is correlated with fan interaction behaviour and video attributes. Firstly, the number of fan reviews is influenced by the number of plays, response rates, the number of likes and time span. From the perspective of fan interaction behaviour, videos that are played more often, liked more often and have a low response rate to a single review are more likely to be reviewed on by subsequent fans. From the perspective of video attributes, videos that have been posted for a long time have a higher number of fan reviews. Secondly, the sentiment of fan reviews is influenced by the time span, video duration and the number of positive reviews. From the perspective of fan interaction behaviour, videos with a high number of positive sentiment reviews have a high average sentiment index of reviews. From the perspective of video attributes, videos with short video duration and long posting time have high average sentiment index of reviews. This finding provides a new perspective for the study of fan reviews in social commerce and enriches and improves the theory of fan reviews and sentiment analysis.

From the perspective of management practice, the findings of the study provide a reference for decision making on both social commerce platforms and the operation of We-media bloggers. Firstly,

from the perspective of increasing the number of fan reviews, social commerce platforms and We-media bloggers should increase the exposure of videos with a high number of plays and likes, and increase the promotion of such videos to help attract more fans to participate in the reviews. At the same time, certain incentives, such as coupons for participation in reviews, can be adopted to promote more positive reviews from fans. Secondly, from the perspective of increasing the emotional intensity of reviews, social media and bloggers should control the videos duration and avoid posting lengthy videos. At the same time, social commerce platforms and bloggers should optimise their review pages by moving positive reviews forward as much as possible in order to drive subsequent fans to post more positive reviews, and by hiding and combining some similar negative reviews in a backward position. This will not only promote the word of mouth spread of the platform and the bloggers themselves, but also allow more fans to quickly see their merits and facilitate their decision making.

Future research could be carried out in two ways. Firstly, the scope of the study should be expanded to include more bloggers from social media platforms to further validate the universality of the paper's findings; secondly, the research methods should be optimised by combining the non-text data in the reviews, such as using deep learning and machine learning, to further explore the influencing factors of fans' behaviour and sentiments and their dynamic change patterns, so as to enrich the management theory of social media.

ACKNOWLEDGEMENTS

REFERENCES

[1] Busalim A H, Ghabban F, Hussin A R C (2021). Customer engagement behaviour on social commerce platforms: An empirical study. Technology in Society 64: 1-17

[2] CNNIC. The 47th Statistical Report on the Development Status of the Internet in China (2021). China Internet Information Center

[3] Internet Society of China. 2021 China Social Commerce Industry Development Report (2021). Available at https://www.sohu.com/a/457713054_120855974. Accessed April 2021

[4] Jin J, Liu Y, Ji P, et al (2016). Understanding big consumer opinion data for market-driven product design. International Journal of Production Research,54: 3019-3041

[5] Li J, Xue W, Yang F, et al (2017). An integrated research framework for effect of EWOM. Journal of Systems Science and Information, 5: 343-355

[6] Li J, Li H (2018). Research on Product Feature Extraction and Sentiment Classification of Short Online Review Based on Deep Learning. Information Studies: Theory and Application, 41:143-148

[7] Zhang H, Zhong H, Hu X (2019). User reviews clustering method based on topic analysis. Computer Science, 46: 50-55

[8] Li J, Lan Q, Liu L, et al (2019). Integrated online consumer preference mining for product improvement with online reviews. Journal of Systems Science and Information, 7: 17-36

[9] Liu H, Zhang X (2017). Research status of E-commerce of our country social analysis based on CNKI literature

research. Journal of Modern Information, 37: 171-177

[10] Zhou T, Chen K (2018). Research on social commerce users' behavior mechanism based on SOR model. Journal of Modern Information, 38: 51-57

[11] Sun Y, Lu Y, Wei K K (2016). Social commerce: The impact of social climate on friendship group members' purchase intention and behavior. Chinese Journal of Management, 13: 1392-1399

[12] Zhou J (2015). User stickiness in the context of Social Commerce: The indirect influence and regulation of user interaction. Management Review, 27: 127-136

[13] Gan C, Zhong Q, Luo T (2017). Factors affecting trust formation in the context of social commerce. Information Science, 35: 68-73

[14] Zhang S, Wei M (2019). The formation of online impulse purchase tendency: An empirical study of the social commerce. Jinan Journal(Philosophy and Social Sciences), 41: 17-29

[15] Shao P, Hu P (2019). Untruthful opinions idenfication and product adoption in consumer advice network. Chinese Journal of Management Science, 27: 96-106

[16] Ding S, Lin J (2020). Factors on consumer continuance intention to share information social media environment: An empirical research from the simultaneous perspective of information seeking and information sharing. Information Science, 38: 131-139

[17] Xiang C, Wang P (2020). Research on multiple adoption behaviors of social media at workplace: Perspective on S-O-R theory. Information Science, 38: 70-76+107

[18] Gan C, Xu J (2020). The effects of perceived value on user behavior intention in the context of social commerce. Information Science, 38: 127-132

[19] Wang A, Zhang Q, Peng Z, et al (2019). Identifying regional demand preferences from online reviews. Chinese Journal of Management Science, 27: 167-176.