Open Access Library Journal
2026, Volume 13, e14908
ISSN Online: 2333-9721
ISSN Print: 2333-9705

An Empirical Study Based on Big Data Analysis to Analyze Social Media Group Sentiment and Its Impact on Consumer Psychology

Xueying Liu

Institute of Artificial Intelligence, Shandong Vocational and Technical University of Engineering, Jinan, China
Email: 1218555700@qq.com

How to cite this paper: Liu, X.Y. (2026) An Empirical Study Based on Big Data Analysis to Analyze Social Media Group Sentiment and Its Impact on Consumer Psychology. *Open Access Library Journal*, **13**: e14908.
https://doi.org/10.4236/oalib.1114908

Received: July 3, 2025
Accepted: February 21, 2026
Published: February 24, 2026

Abstract

As social media becomes the core domain of information interaction in the era of big data, the emotional information contained in the vast amount of user-generated content provides an unprecedented data foundation for understanding group psychology and behavioral laws. Based on the intersection of big data engineering and psychology, this study aims to construct a comprehensive analytical framework to explore the law of group mood swings in social media and its predictive effects on consumer psychology and behavior. This study comprehensively uses natural language processing, machine learning and time series analysis to comprehensively calculate and model sentiment in multi-platform social media texts, and constructs a consumer psychology prediction model based on this. Empirical analysis reveals the evolution patterns of group sentiment under the influence of time, event-driven, and social network structures, and confirms the critical role of sentiment traits in predicting consumer behavior. This study not only enriches the theoretical enlightenment of computational psychology and consumer behavior, but also provides data-based decision support and practical paths for enterprise precision marketing, public sentiment governance, and personal consumption decision-making.

Subject Areas

Artificial Intelligence, Big Data Search and Mining

Keywords

Big Data Analysis, Social Media, Group Emotion, Emotional Infection, Consumer Psychology Prediction, Natural Language Processing

1. Introduction

1.1. Background of the Study

We have fully entered a new era driven by data, where the explosion of social media has become a core component of the big data ecosystem. According to statistics, there are more than a billion active social media users in China alone, generating hundreds of terabytes of unstructured data such as text, images, and videos every day. This data not only records users' social interactions and life trajectories, but also maps the public's collective psychological state, mood swings, and potential behavioral intentions. In this context, how to automatically and intelligently extract and understand group sentiment information from large, real-time, and heterogeneous social media data, and explore its influence mechanism on economic and social behavior, especially consumer behavior, has become a cutting-edge topic of common concern in many disciplines such as big data science, psychology and management.

The combination of psychology and big data has given rise to the emerging field of "computational psychology", which leverages data science techniques such as natural language processing (NLP), machine learning, and complex network analysis to achieve quantitative measurement, dynamic tracking, and modeling and prediction of human mental states. Group emotion, as an important concept in social psychology, focuses on the emotional experience shared by group members in specific social situations and their dynamic transmission process [1]. With their openness, immediacy, and connectivity, social media platforms provide a near-perfect "digital field" for observing group sentiment at scale and in its natural state. At the same time, consumer psychology and behavioral prediction have always been at the heart of corporate marketing and business intelligence. Traditional research methods have limitations such as high cost, limited samples, and lags in timeliness. Social media-based data analysis can capture early signals of consumer psychology in real time, covering multiple dimensions such as user emotional expression, social interaction, and content preferences, enabling more accurate and forward-looking predictions.

Therefore, this study focuses on two closely related topics: "social media group sentiment analysis" and "consumer psychological impact prediction", aiming to construct a full-chain analysis framework from data perception to decision support, through big data engineering technology, and deepen the understanding of the complex relationship between emotion and consumption in the digital age.

1.2. Research Status at Home and Abroad

Scholars at home and abroad have conducted extensive research on social media sentiment analysis and its applications. In terms of group sentiment analysis, early studies such as Gold and Messi used Twitter data to reveal circadian and seasonal rhythms of mood changes [2]. Based on Weibo data, Wang *et al.* described in detail the daily, weekly, and seasonal mood swings of Chinese netizens, and found that positive emotions peaked at noon and evening, negative emotions decreased significantly on weekends, and mood swings were strongest in summer and calmer in autumn [3]. These studies confirm the feasibility of macro sentiment measurement based on social media big data. In terms of emotional transmission mechanisms, Gao *et al.* systematically reviewed social media emotional infection models, pointing out that epidemiological dynamics (such as SIR and its variants) and information cascade models are two mainstream modeling paradigms, revealing how emotions diffuse and evolve through user interaction in social networks [4].

In terms of consumer psychology prediction, research focuses on using direct feedback data, such as online reviews and ratings, to predict satisfaction or purchase intent. However, macro- or mesoscale group emotional states have not been fully explored as leading indicators for predicting consumer behavior. There are three shortcomings in existing research: first, most sentiment analysis studies only focus on the description and dissemination simulation of emotions, and fail to establish a predictive correlation with downstream economic behavior (especially consumption). Secondly, the dimension of sentiment analysis is relatively single, usually limited to positive and negative binary classifications, lacking detailed descriptions of discrete emotions or emotional dimensions (value, arousal), and different specific emotions (such as anger, sadness, joy) may have different effects on consumption. Third, the psychological influence mechanism of emotions on consumption is insufficient, and most of them are still in relevant analysis, and there is a lack of in-depth research on their media or moderating variables.

1.3. Purpose and Significance of the Study

This study aims to fill the above research gaps with the following specific objectives:

1) Establish a multi-source social media data fusion collection and preprocessing process to ensure the scale, quality, and timeliness of data.

2) Develop a multi-level sentiment analysis model suitable for the Chinese social media background, and realize comprehensive sentiment analysis from binary and multi-category discrete emotions to continuous dimensions (value, arousal).

3) The temporal fluctuation law of group emotion is systematically analyzed, its relationship with the calendar cycle (day, week, season) and major social events is discussed, and the infection and propagation mode of emotion in the network structure is preliminarily discussed through complex network theory.

4) A consumer psychology prediction model is constructed, integrating emotional characteristics, user behavior characteristics, and background characteristics, and quantitatively evaluating the intensity and pattern of the influence of group and individual emotions on consumer intentions or behaviors.

5) Based on the research results, targeted recommendations are made from the perspectives of corporate marketing, public governance and personal health networks.

This study has theoretical value and practical significance. At the theoretical level, it promotes the in-depth intersection of psycho-emotional theory, social communication theory and big data computing methods, provides a new methodological framework for the study of "digital society mentality", and empirically tests the complex relationship model between emotion and consumer behavior. At the practical level, the findings can be applied to: providing enterprises with real-time market sentiment dashboards and consumer trend alert systems to help precision marketing and product strategy optimization; provide government departments with online public opinion and public opinion monitoring tools to help prevent social risks and maintain harmony and stability. It provides a reference for individual users to understand their own emotions and rational consumption guidance.

2. Theoretical Basis and Analytical Framework

2.1. Core Theoretical Structure

This study focuses on three theoretical pillars:

1) Psycho-emotional theory: includes Ekman's discrete basic emotion theory (happiness, sadness, anger, fear, disgust, surprise) [5] and Russell's cyclic emotion model (value-evocation two-dimensional model) [6]. The former helps visually label the mood of social media text, while the latter better captures the intensity and mix of emotions. In addition, Mackie *et al.*'s group emotion theory emphasizes that emotions on social media are not only the convergence of individual emotions, but also the "collective psychological reality" formed through social comparison, group identity, and emotional infection [1].

2) Emotional Infection Theory: Emotional infection refers to the process of unconscious transmission of emotional states between individuals or groups. In social media, emotions are transmitted through a network of users through information carriers (posts, comments, retweets). Based on the review by Gao Xiaoyuan *et al.* [4], this study believes that the emotional infection process is influenced by network structure, user relationship strength, information characteristics, and individual differences, which are key mechanisms for understanding the formation and evolution of group emotions.

3) Consumer Behavior Theory: Kahnemann's prospect theory states that decisions often deviate from full rationality and are influenced by emotional and cognitive frameworks [7]. Consumer psychology research also suggests that positive emotions may promote impulsive purchases and hedonistic spending, while negative emotions may trigger compensatory spending (such as "shopping therapy") or discourage willingness to spend. The group emotional atmosphere on social media may affect the emotional state of individual consumers through emotional infection mechanisms, thereby interfering with their decision-making process.

2.2. Technical Analysis Framework

This study adopts a three-layer technical framework of "data layer, analysis layer and application layer".

Data layer: Responsible for collecting multimodal data (mainly text) from China's representative social platforms such as Weibo, Douban, and Zhihu, and performing preprocessing such as cleaning, denoising, word segmentation, and emotional word recognition. Particular attention is paid to the emotional encoding of web words and emojis.

Analytics layer: This is the core layer and contains two submodules.

Group Sentiment Analysis Module: Utilize pre-trained language models (e.g., BERT, RoBERTa) to fine-tune and build multi-layer sentiment analysis models. then time series analysis (such as seasonal decomposition and event research methods) is performed to reveal the law of mood swings; The emotional communication network is built based on the user's attention/forwarding relationship and analyzes the emotional infection path.

Consumer Psychology Prediction Module: Based on the results of sentiment analysis, a comprehensive feature system is constructed, including user emotional history, real-time emotional state, emotional infection index, participation in consumer topics, social network attributes, etc. Predictive models are built using ensemble learning algorithms such as LightGBM and XGBoost, using SHAP and other methods to analyze the interpretability of the models to clarify the contribution of each feature, especially the sentiment profile.

Application layer: Turn analytics into actionable insights and output policy recommendations for different audiences (businesses, governments, individuals).

3. Research Design and Methods

3.1. Data Collection and Processing

In order to comprehensively explore the relationship between group sentiment and social media consumer psychology, this study adopts a multi-platform data fusion strategy, and selects three representative social platforms in China, Weibo, Doubian and Zhihu as data sources. The data collection covers 18 months from January 2023 to June 2024, aiming to cover the entire annual cycle and diverse networking activities.

In terms of data collection, research focuses on public content that is highly relevant to consumer behavior and emotional expression. Weibo data mainly combines an open interface and targeted collection methods to obtain original blog posts and interactive comments around core tags such as consumption, shopping, emotion, and psychology. Doubian Data has selected panel discussion posts that include reflective consumption concepts, such as "The Return of Consumerism" and "Is Consumption Falling Today?" Zhihu Data mainly collects high-quality Q&A related to consumer psychology and emotional management. Finally, this study constructed a multi-source dataset containing approximately 1.9 million valid texts, strictly adhering to academic ethics, collecting only publicly available information and anonymizing user identities to ensure research compliance.

In the face of noise and non-canonical expression in social media texts, a systematic data preprocessing process is developed. First, perform data cleansing to remove ad content, duplicate text, and irrelevant content; The text is then normalized, including uniform character encoding, standardized date and number formatting, and the conversion of emojis into emotional semantic labels. Subsequently, Internet buzzwords were standardized and mapped. Finally, Chinese vocabulary segmentation and keyword extraction are carried out, and general obstacle words are filtered. Through this process, raw unstructured text data is transformed into high-quality structured datasets that can be used for in-depth analysis.

3.2. Construction of Group Sentiment Analysis Model

In terms of sentiment analysis, this study adopts a multi-level analysis framework. The final method of determining sentiment annotation is mainly based on fine-tuning of pre-trained Chinese BERT models such as bert-base-Chinese. We build multi-task learning or hierarchical models to make synchronous predictions:

1) emotional polarity (positive/negative/neutral);

2) discrete emotion categories (based on Ekman theory, such as joy, sadness, anger, etc.);

3) continuous valence and arousal scores (based on the Russell model, ranged from −1 to 1 and 0 to 1, respectively).

For model training and evaluation, we manually annotated 20,000 texts (stratified sampling by platform and time) as the gold standard. The average Krippendorff's α coefficient was higher than 0.75 in the categories of emotional polarity and discrete emotion. The valence and arousal scores are the average of the annotator's scores. The dataset is divided into training set, validation set and test set according to the ratio of 7:1.5:1.5. Hyperparameter tuning (e.g., learning rate at (1e−5, 5e−5), batch size at (16, 32) and early stop were performed using validation sets.

To aggregate individual text sentiment at the daily group level, we calculate the daily "group sentiment index". For day d, its group sentiment index GSI_d is defined as:

$$GSI_d = \frac{\sum_{i \in T_d} \left(w_i \cdot \left(V_i + A_i \right) \right)}{\sum_{i \in T_d} w_i}$$

Among them, T_d is the collection of texts published on the dth day, V_i and A_i are the valence and arousal scores of text *i*, respectively, and w_i is the weight factor that characterizes the potential influence of the text, and the calculation formula is $w_i = \log\left(1 + \text{likes}_i + \text{retweets}_i\right)$. This index synthesizes the emotional tone of a text and its potential dissemination within the platform.

3.3. Emotional Infection Construction and Consumer Psychology Prediction Model

In terms of emotion transmission mechanism, this study analyzes the emotion diffusion process based on the social interaction relationship between users. We built a directed network with nodes as users and edges representing forward/comment/reply relationships. Each edge is labeled with the emotional polarity of the source content. Network metrics (e.g., clustering coefficients for different sentiment cascades, average path length) were calculated to describe the infection pattern.

In terms of consumer psychology prediction, this study constructs a multi-dimensional feature system, including user sentiment characteristics (such as recent average sentiment value, exposure to GSI on the day), consumption-related content characteristics (such as consumption topic keyword frequency), behavior pattern characteristics (such as posting frequency, social centrality), and background attribute characteristics (such as platform, profile-based inferred gender/age). The target variable of the prediction task is a binary tag that indicates whether the user has shown spending intent in the short term (next 7 days), which is obtained by analyzing posts posted by users containing clear purchase plans, product consultations, or the use of shopping-related hashtags, combined with keyword matching and manual sampling verification.

We use the LightGBM algorithm for prediction. The model is trained on the training set using 50% fold cross-validation. Bayesian optimization is used to optimize the hyperparameters such as the number of leaf nodes, learning rate, and feature sampling ratio. Model performance was evaluated using accuracy, precision, recall, F1 score, and area under the ROC curve (AUC). The final model performed well, with an accuracy of 0.816 and an AUC of 0.878. To interpret the model, SHAP (SHAPLEY Additive exPlanations) was used to analyze the directional influence of feature importance and key sentiment features on the prediction results.

4. Results and Analysis

4.1. Data Description

After preprocessing, the final dataset contains 1.92 million valid texts. The data mainly comes from Weibo (about 1.45 million, accounting for 75.5%), Douban (about 280,000, accounting for 14.6%) and Zhihu (about 190,000, accounting for 9.9%), covering various types such as real-time discussions, in-depth community exchanges, and knowledge Q&A.

4.2. Sentiment Analysis Model Performance

The fine-tuned BERT model performed well on the manual annotation test set. The macro average F1 score of the three categories of emotional polarity was 0.842, and the macro average F1 score of the six categories of discrete emotions was 0.713. The root mean square error (RMSE) of potency prediction is 0.186, and the RMSE of arousal prediction is 0.211, indicating that the model can effectively capture fine-grained emotional information of text.

4.3. Temporal Patterns of Group Sentiment

The calculated daily Group Sentiment Index (GSI) shows clear fluctuations and event-driven peaks and valleys over 18 months. A deeper temporal analysis revealed systematic patterns: the GSI (more positive/arousal) was significantly higher on weekends than on weekdays, with the lowest median on Mondays. The median GSI was highest and the variance was highest in summer, and the lowest was in autumn, indicating calmer or slightly depressed collective mood.

4.4. Analysis of Emotional Infection Networks

Analysis of the constructed sentiment-labeled forwarding networks showed that the negative emotion cascade propagated more hops than positive sentiment on average and formed tighter local subgroups, which supports the finding that negative emotions may be more contagious or persistent in some network structures.

4.5. Results of Consumer Psychology Prediction Model

The LightGBM model has achieved solid performance in predicting short-term consumption intentions. On the independent test set, the model accuracy is 0.816, the accuracy is 0.793, the recall is 0.754, the F1 score is 0.773, and the AUC is 0.878.

SHAP analysis showed that characteristics associated with direct participation in consumption topics were the strongest predictors. Sentiment characteristics—including the user's recent average valence, exposure to the current day's GSI, and emotional intensity (arousal) of their own posts—are among the top 10 important features, confirming the significant predictive value of sentiment signals.

To understand the nuanced correlation between sentiment and consumption predictions, we generate bias maps for key sentiment characteristics. The biased dependence graph of the average user valence graph is inverted U-shaped: the predicted probability of consumption intention increases with the increase of potency until it reaches a moderately positive level (about 0.4 - 0.6), and then decreases slightly at the high efficiency value. The biased dependence graph of GSI exposure shows a positive but saturated association, indicating that being in a positive group atmosphere can help increase consumption intention, but there is a diminishing return effect at very high GSI levels.

5. Conclusion and Discussion

5.1. Key Findings

Through the empirical analysis of big data, the following core conclusions are drawn:

1) Social media group sentiment shows regular fluctuations and event sensitivity. Its changes are related to factors such as biological rhythms, social routines, seasonal climate, and major social events, and negative emotions are more contagious in social networks [4] [8].

2) Group and individual emotions are effective predictors of consumer psychology. The constructed prediction model confirms the key role of emotional factors in predicting consumption behavior, and its contribution is significant. The AUC of the model reached 0.878, and the importance of emotional characteristics ranked high.

3) The relationship between emotion and consumption is complex and nonlinear. Its performance is not simply "positive promotion and negative inhibition", but also involves complex mechanisms such as "optimal emotional interval" and "emotional repair". This nonlinear mode is intuitively supported by the dependent graph. Emotional infection may play an important regulatory role in this.

4) The technical framework constructed by the institute has interdisciplinary methodological value, effectively integrating psychological concepts, communication mechanisms and data science tools, and providing a feasible path for quantitative research on psychosocial psychology and behavior in the digital environment.

5.2. Theoretical Contributions and Practical Enlightenment

Theoretical contributions:

This paper deepens the theoretical understanding of the relationship between emotion and consumption, and promotes the empirical test from simple linear correlation to nonlinear and conditional effects. The association between the emotional atmosphere at the group level and the consumption decision at the individual level provides a basis for micro-macro cross-level analysis. The application of emotional infection theory in the field of consumption is verified and expanded [4], and the potential influence path of "emotional infection → individual emotions → consumption decisions" is proposed.

Practical enlightenment:

For corporate marketing, it is recommended to establish a "social sentiment monitoring system" to identify group sentiment cycles and event windows; Strengthen marketing during periods of moderately positive public sentiment; pay attention to user groups with high emotional appeal; Understand the differentiated needs of consumers in different emotional situations (such as "comfort" products in negative times). For public governance, government departments can use similar models to conduct social mentality early warning, take the initiative to carry out information counseling and psychological intervention during the period of negative emotional accumulation (such as after major emergencies), prevent irrational collective behavior, and pay attention to guiding the concept of healthy consumption. For individual users, enhance the meta-cognition of the relationship between their own emotional state and consumption decisions, consciously delay making major consumption decisions in extreme emotional states (whether positive or negative), and cultivate rational consumption habits.

5.3. Research Limitations and Future Prospects

Limitations:

1) Data ecological validity: It mainly relies on public text data, making it diffi-

cult to obtain private social media (such as WeChat Moments) and real transaction data, and there is sample bias.

2) Causal inference challenges: Despite the use of predictive models and event research methods, strictly establishing the causal relationship between emotions and consumption still requires more complex experimental or natural experimental design.

3) Complexity of sentiment measurement: Text sentiment analysis is difficult to fully capture complex expressions such as sarcasm and obscurity, and lacks multimodal signals such as physiology and speech.

Future Outlook:

1) Multimodal fusion analysis: Integrate images, emojis, and short video content to build a more powerful emotion recognition system.

2) Dynamic network and evolutionary modeling: Combined with time series graph neural network, the collaborative communication process of emotion and consumption intention in dynamic evolutionary social networks is simulated.

3) Cross-cultural comparative research: Collect social media data from different languages and cultural backgrounds to explore the cultural specificity and universality of emotion-consumption relationships.

4) Privacy protection and ethical norms: With the popularization of affective computing technology, there is an urgent need to formulate relevant data ethics guidelines and user privacy protection frameworks to ensure that technology is good.

6. Epilogue

Big data opens a new window for human social psychology. Through technological integration and empirical exploration, this study preliminarily reveals the complex picture of group sentiment analysis and its impact on social media consumer psychology. In the future, with the continuous advancement of technology and the deepening of interdisciplinary collaboration, research in this field will not only more accurately depict the "socio-emotional pulse" of the digital age, but also provide smarter data drivers for promoting rational consumption, healthy markets, and good social attitudes.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Mackie, D.M. and Smith, E. R. (2015) Intergroup Emotions Theory: Production, Regulation, and Modification of Group-Based Emotions. *Advances in Experimental Social Psychology*, **58**, 1-69.

[2] Golder, S.A. and Macy, M.W. (2011) Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength across Diverse Cultures. *Science*, **333**, 1878-1881. https://doi.org/10.1126/science.1202775

[3] Wang, J.Y., Gan, S.Q., Zhao, N., *et al.* (2016) Analysis of Emotional Change Based on

Weibo Users. *Journal of Chinese Academy of Sciences*, **33**, 815-824.

[4] Gao, X.Y., Liu, Z., Chai, Y.J., *et al*. (2019) Research on Social Media Emotional Contagion Model. *Applied Psychology*, **25**, 372-384.

[5] Ekman, P. (1992) An Argument for Basic Emotions. *Cognition and Emotion*, **6**, 169-200. https://doi.org/10.1080/02699939208411068

[6] Russell, J.A. (1980) A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, **39**, 1161-1178. https://doi.org/10.1037/h0077714

[7] Kahneman, D. and Tversky, A. (1979) Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, **47**, Article 263. https://doi.org/10.2307/1914185

[8] Fan, R., Zhao, J., Chen, Y. and Xu, K. (2014) Anger Is More Influential than Joy: Sentiment Correlation in Weibo. *PLOS ONE*, **9**, e110184. https://doi.org/10.1371/journal.pone.0110184