

# A study on predictive modeling of users' parasocial relationship types based on social media text big data

Jiatong Meng\*,  
Universidad Carlos III de Madrid, Madrid,  
Spain

Yucheng Chen  
Software College, Northeastern University,  
Shenyang, 110000  
China

\*Email of corresponding author: jtm921@sina.com

Received: May 14, 2021. Revised: December 17, 2021. Accepted: January 6, 2022. Published: January 7, 2022.

**Abstract**—The traditional parasocial relationship type prediction model obtains prediction results by analyzing and clustering the direct data. The prediction results are easily disturbed by noisy data, and the problems of low processing efficiency and accuracy of the traditional prediction model gradually appear as the amount of user data increases. To address the above problems, the research constructs a prediction model of user parasocial relationship type based on social media text big data. After pre-processing the collected social media text big data, the interference data that affect the accuracy of non-model prediction are removed. The interaction information in the text data is mined based on the principle of similarity calculation, and semantic analysis and sentiment annotation are performed on the information content. On the basis of BP neural network, we construct a prediction model of user's parasocial relationship type. The performance test data of the model shows that the average prediction accuracy of the constructed model is 89.84%, and the model has low time complexity and higher processing efficiency, which is better than other traditional

models.

**Keywords**—Social Media, Textual Big Data, Parasocial Relationship, Relationship Type Prediction, Prediction Model, Neural Network

## I. INTRODUCTION

THE rapid development of social media (such as livestreaming, Douyin, Vlog, etc.) has prompted scholars to refocus on the theory of parasocial interaction. Different from traditional media, social media provides new research contents and ideas for scholars to explore the parasocial interaction between media figures and consumers due to its characteristics of instantaneity of communication, interaction of modes, novelty of content, individuation of media and scene of display. Social network platforms have attracted a large number of active users, and the social relationships among users have formed an intricate network graph. How to tap the potential information in social network platforms has become a hot research topic nowadays. One of the important research directions is the prediction of social media users'

parasocial relationship types. It inferred the type of parasocial relationships among social media users by evaluating information such as the strength of potential relationships among them. Parasocial relationships initially refer to the kind of realistic social relationships formed between mass media audiences and the representations of people appearing in the media, similar to those formed by people through daily face-to-face communication. That is, the audience obtains the words and actions of a celebrity through the media, becomes emotionally attached to him or her, and develops an imaginary interpersonal relationship based on this. Unlike two-way, real social interactions, in this social relationship, the audience is unilaterally familiar with the media figure and accepts its influence, but the media figure can only guess the audience's behavior and does not meet the definition of a relationship in the true sense, so it is called a parasocial relationship [1], [2].

It was found that the parasocial relationship between the audience and the media figure is a hypothetical intimate relationship between the audience and the media figure constructed through the audience's imagination. Initially, the parasocial relationship was considered to be completely one-way, with the initiative in the hands of the mediator, and the audience had no control over the existence and development of the relationship. Further research revealed that social media users subjectively feel an interaction with the mediator and sense that the mediator's behavior is influenced by their own feedback, even though they know it is an illusion, such as the mediator only pretending to know the audience's reaction because they only act on the basis of the expected reaction, not on the observed reaction. At the time when the audience does not feel that they have a relationship with the media figure, a parasocial relationship has been created, the intensity and breadth of which will be related to factors such as the frequency and persistence of the media figure observed by the audience. In contrast to the state of parasocial relationships in traditional media, users on social media in the Internet era generally have more intense parasocial relationships [3]. The growth in the volume of social media users has contributed to the formation of parasocial relationships among media users. By predicting the types of parasocial relationships among social media users, it is possible to conduct an in-depth study of the effects of

information dissemination in social media to make up for the long-standing lack of research in this field from the perspective of user relationships. The model for predicting the type of parasocial relationship among users studied in the literature [4], [5] automatically acquires target data based on set keywords using crawlers and other means, and obtains the prediction results of the type of parasocial relationship through the frequency of communication among users. The model has good prediction accuracy for traditional media with poor user interaction, but the number of modern social media users is large and the content of social information is complex, so there is a large uncertainty in judging the parasocial relationship of users by keywords alone, which leads to a large error in the prediction results of the prediction model. The prognostic model in the literature [6] uses a combination of discrete GM (1, 1) and gray prediction principles to obtain the corresponding prognostic results through the function transformation and inverse transformation processing of the research data. This model has a complex and cumbersome processing process, a limited scope of application with low prediction efficiency and low prediction accuracy. Literature [7] was proposed based on friend relationship prediction information recommendation algorithm and the user's information recommended in the prediction process, social relations are based on topology information and user's mutual information based on the user and the existence of social relationships determine relationship type determination, and by using linear regression analysis and logistic regression analysis method to realize the integration of basic characteristics.

In order to solve the problems of low prediction accuracy and low processing efficiency of the traditional parasocial relationship prediction model and its inability to adapt to the current network social media environment, this paper introduces a new theory to build a new prediction model. Currently, social media users still interact with each other on social media software in the form of textual communication, and the textual content used in user interaction is easier to analyze users' psychological and other detailed information when they are socializing. Therefore, compared to video and audio forms, the text data used by social media users when communicating can more comprehensively analyze the type of

parasocial relationships of users. Therefore, based on the above analysis content, this paper will establish a prediction model of user parasocial relationship types by analyzing and processing social media text big data.

## II. A STUDY ON PREDICTIVE MODELING OF USERS' PARASOCIAL RELATIONSHIP TYPES BASED ON SOCIAL MEDIA TEXT BIG DATA

### A. Social media text big data pre-processing

Before constructing a predictive model of user parasocial relationship types, user social text data within the scope of the study requirements were crawled from the database of the social media object platform selected for the study, as well as using crawler technology. The data is stored in TXT file in string format, and the string content in TXT is first read out. Since the background storage database of social media is a pre-processed text database, its data format and some word information are not what we want, and Python's regular expression Re library can be well into filtering the string type data, so the following processing of the obtained word database is made using regular expressions [8] - [10].

(1) Removal of punctuation: the text data in the corpus is represented in the form of a sentence, where punctuation is not needed to create a corresponding lexicon, so by regular expressions, punctuation is removed and replaced by spaces. Periods are replaced by # signs, which are used to cut different sentences.

(2) Remove the date: each user's social message in the backend storage database of social media is preceded by a date marker, by removing this to ensure the purity of the user's social content.

(3) Read relevant markers: Each word unit in a user's social content is marked, and social media users of different age groups use social media's own or expanded emoji icons when socializing. The different emoticons represent the user's attitude and even the content of the communication. In addition, different age groups of social media users have different understanding and usage of the same emoji icons, which need to be differentiated when reading the tokens. We extracted all the word tags and stored them in an array, and removed the tags for general punctuation.

After processing the users' communication data on the

corresponding platforms, the interference data generated by inactive users is filtered to improve the accuracy of the subsequent model construction.

Inactive users are those who are not real users, but those who are maliciously registered by machines or software or those who have very low activity over time. These users cause some confusion to the predictive study of social media users' parasocial relationships and must be eliminated beforehand.

There are three main types of methods for detecting inactive users.

(1) Detection of inactive users from the user base using user characteristics such as "number of followers", "number of followers" and "number of social media posts and interactions".

(2) The nicknames and tweets posted by inactive users have obvious traces of random machine generation, which can be used to detect inactive users.

(3) Inactive users usually follow a real user in order to increase the number of fake followers of the real user, and thus the inactive users have a more special following relationship, and the inactive users can be detected by mining the user relationship. Among the above three types of methods, the latter two algorithms have larger time complexity and are not suitable for use in the pre-processing stage of data.

In this section, we propose a simple algorithm for detecting inactive users by combining user behavior characteristics. The facile algorithm first defines the rate of change of social media posts [11], [12]:

$$m'_i(t_m, t_n) = \frac{m_i(t_n) - m_i(t_m)}{t_n - t_m} \quad (1)$$

In the above equation,  $t_m$  denotes the start time;  $t_n$  indicates the end time.  $m_i(t_n) - m_i(t_m)$  indicates the number of interactions posted on social media by user  $i$  during the time period  $t_m \cdot t_n$ . A two-month statistical cycle is used. For each user, the rate of change of posting interaction information on social media in the most recent year is calculated. Then the user's activity on social media is defined by the following equation:

$$act.T(i) = \frac{\max_{1 \leq k \leq 6} m'_i(t_m, t_n)}{\frac{1}{N} \sum_{j \in V} m'_i(t_m, t_n)} \quad (2)$$

In the above equation, the numerator denotes the maximum value of the social media change rate of user  $i$  in the most recent year; the denominator denotes the average interactive posting change rate of all users in the social media network;  $V$  denotes all users of the corresponding social media, and  $N$  denotes the number of all users. The activity of users' attention can be defined according to the following equation:

$$act\_F(i) = \frac{f_o(i)}{\frac{1}{N} \sum_{j \in V} f_o(i)} \quad (3)$$

In the above equation,  $f_o(i)$  denotes the number of followers of user  $i$  and  $V$  denotes all users within the microblogging network. Then, the user's activity is defined as follows:

$$ACT(i) = act.T(i) * act\_F(i) \quad (4)$$

The determination condition of inactive users is  $ACT(i) < \alpha$ . Where  $\alpha$  is the threshold value, and after several trials, the threshold value is taken as 0.01 with the best effect, which can remove all the data of inactive users on social media. After the above pre-processing of social media text big data, the similarity principle is used to mine the text big data information for predicting the type of parasocial relationship of users.

#### B. Social media text big data information mining

A HowNet-based word similarity calculation method was chosen to mine the information in social media text big data. HowNet specifies the use of one or more concepts to describe a word, each concept consists of a series of sense-primitives, and the similarity formula between sense-primitives is as follows [13]:

$$sim(s_i, s_j) = 0.5 * \frac{l}{h+d} + 0.5 * \frac{2 * \log p(C)}{\log p(s_i) + \log p(s_j)}$$

In the above equation,  $l$  denotes the number of levels where the least common node of the righteous original  $i$  and the righteous original  $j$  is located;  $h$  denotes the depth of the righteous original tree;  $d$  denotes the distance between the righteous original  $i$  and the righteous original  $j$ ;

$p(s)$  denotes the ratio of the number of nodes directly connected to the righteous original  $s$  to the number of summary points of the righteous original tree, and  $C$  denotes the least common parent node of the righteous original  $i$  and the righteous original  $j$ . Using  $n$  to denote the number of sense-primitives and  $sim_k$  to denote the similarity of the  $k$ th sense-primitive; denoting the normalized weight, the formula for the similarity between concepts is as follows [14], [15]:

$$sim(c_i, c_j) = \sum_{k=1}^n a_k sim_k \quad (6)$$

The semantics in the text information is analyzed according to the above-mentioned results of similarity calculation between word concepts. A three-layer Bayesian structure is used as the core for semantic analysis of text, subject, and word in turn. A certain subject is selected according to a certain probability distribution, and then a certain word is selected from this selected subject according to a certain probability distribution, and so on, until the corresponding semantic analysis results are obtained.

After getting the semantics of the text when users post corresponding social messages, we need to further analyze the sentiment of the text messages posted by users. In addition to the semantics of the text content used by social media users during interaction, the symbols such as emoticons used by users during communication and interaction also contain the emotions of users during social interaction. Analyzing the textual sentiment of users during social interactions can assist in the prediction of users' parasocial relationships.

#### C. Textual big data sentiment analysis

When performing sentiment analysis of text data, the research aims to tag social texts with sentiment. Sentiment tagging is divided into two categories of positive and negative sentiments. So the support vector machine was chosen to be used as the training algorithm for sentiment analysis. Through effective training and testing, the algorithm has good classification results in the process of sentiment classification.

Classification can be achieved using the optimal classification hyperplane when the data samples can be separated in a linear space, and SVM can solve this problem when the data samples do not satisfy linear separability. As shown in Fig. 1, if asked to manually animate a classification

line to distinguish between hollow and solid circles, most people would prefer to draw a line close to the position on the line AB because, this line is farthest from both samples. The basic idea of the optimal classification hyperplane in the two-dimensional case is also revealed in Fig.1 The optimal classification hyperplane needs to satisfy that the data samples can be separated correctly, while those sample points of the two categories of data samples that are closest to the hyperplane have the greatest distance between them and the hyperplane [16], [17].

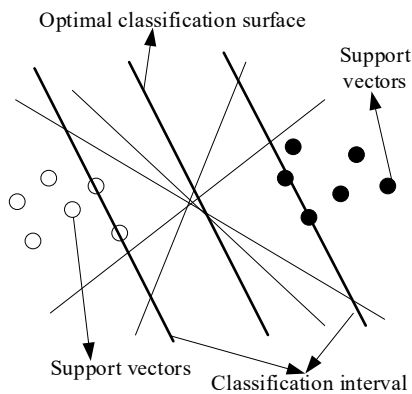


Fig. 1 Two-dimensional diagram of the optimal classification surface

For this linear classifier of this supervised learning algorithm, the expression of each classification score function for the evaluation data can be expressed as [18]:

$$f(x, W, b) = Wx + b \quad (7)$$

In the above equation,  $x$  is the input data matrix for SVM sentiment analysis;  $W$  is positive or negative sentiment; and  $b$  is a parameter. The corpus of interactive text with sentiment markers is put through correlation processing to obtain the words and associated sentiment markers in it. Then the word text is used as input data for the Word2vec toolkit mentioned above, and the interactive sentiment text sentiment word vector database is obtained by training, and each word vector is noted as  $V'$  [19]. Since the basic unit of sentiment markers needed in sentiment analysis is the sentence, it was chosen to take the word vectors of all words in the sentence to be accumulated and then averaged to represent the feature vectors of the sentence. The obtained sentiment markers are divided into positive and negative categories, and 0, 1 numbers are used to represent positive and negative sentiment

respectively. Based on the above detailed analysis and processing of the social media text big data, a BP neural network is used as the basis for the construction of the prediction model to establish the user parasocial relationship prediction model.

#### D. Complete the construction of a predictive model of parasymphathetic relationship types

The model of BP neural network consists of 3 parts, which are input layer, implicit layer and output layer. The input layer provides a bridge for human-computer interaction, and its main task is to receive the information from the outside world and pass it to the implicit layer; the implicit layer is the core of the BP neural network, and its job is to process the data and information passed by the implicit layer; the output layer is the final processing of the information passed by the last implicit layer, and then the output of the processing result [20].

In general, there are many implicit layers in BP network. After several weight adjustment, the network can learn sample knowledge and store it in the form of weight. However, there are too many implicit layers, the number of weights that need to be adjusted increases greatly, and network learning takes up more time. Theoretically, when the number of implicit layer neurons (nodes) can be arbitrarily set, an implicit layer three-layer BP neural network can approach any continuous function with arbitrary precision. Therefore, in most applications, only one implicit layer of three-layer BP neural network is used. The structure of BP neural network is shown in Fig. 2.

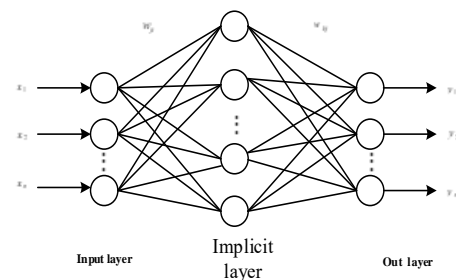


Fig. 2 BP network model

The activation function expression of the BP neural network is set in this study as follows:

$$R(x) = e^{-|x|^2} \quad (8)$$

For the three-layer BP neural network, the input layer

outputs the processed data to the implicit layer under the action of activation function after information mining, semantic analysis and sentiment labeling of the input data. The data in the implicit layer classifies the input data according to the principle of clustering, and labels the classification labels

on the corresponding data and outputs them to the output layer, which outputs the corresponding results by the output layer of the BP neural network. The simulation flowchart for discriminating user parasocial relationship types using BP neural network algorithm is shown in Fig. 3 [21].

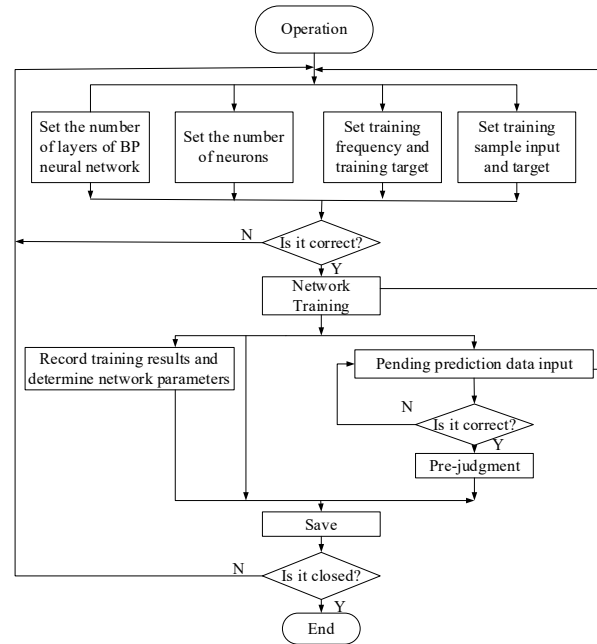


Fig. 3 Flow chart of BP network algorithm for predicting user's parasocial relationship type

After the initial parameters of the neural network are set to 0, the sample data set is used to train the network parameters and determine the final prediction model parameters. After the data to be processed are input to the BP neural network with determined parameters, the classification result output by the neural network processing is the user's parasocial relationship type judged by the prediction model. Thus, the construction of the prediction model of user parasocial relationship type based on social media text big data is completed [22], [23].

### III. MODEL PERFORMANCE TESTING STUDY

The prediction model of user parasocial relationship types based on social media text big data was constructed above, and this section conducts an experimental study based on real

social media interaction datasets to prove that the prediction model of parasocial relationship types constructed in the above study is effective.

#### A. Experimental data preparation

The dataset selected for this experiment is a collection of more than 5000 users and 60 million text-based social interactions collected from May 1, 2020 to December 31, 2020 using crawlers on four social media: Weibo, Post, Zhihu, and Douban. All the information on users and social media are represented in TXT format, and Table I gives the list of some attributes used in this experiment.

Table. I User and interactive field attributes used in this experiment

Entity	Fields	Description
User	User UID	Unique identifier for the user
	User nickname	User's screen name
	Number of fans	Follow the number of users of a user

Social media interactive information	Number of friends	Number of followers of a user
	Number of satus	Number of all social interaction messages posted by this user
	Personal Description	Free text in personal data
	Location	Free text of the user's location
	Text	Content of socially interactive text, including words, interactive objects, tags, etc.
	Timestamp	Time of posting (or retweeting) of socially interactive texts
	Original text	Content of raw social interaction text data

The experimental data were divided into data sets with the same amount of data and tested for multiple experiments. All the above data were processed according to the processing requirements of the corresponding models before the experiments to avoid interference with the experimental results.

#### B. Experimental content

The experiments were conducted in the form of comparison, with the prediction model of user parasocial relationship types based on social media text big data constructed above as the experimental group, and the traditional parasocial relationship prediction models mentioned in literature [4] and literature [5] as the comparison groups I and II. The accuracy and

efficiency of the three parasocial relationship prediction models are compared from the perspectives of prediction accuracy and prediction processing time complexity of the prediction models, respectively.

#### C. Experimental results and analysis

Three sets of parasocial relationship prediction models are used to predetermine the experimental data set, and all models have the same criteria for predetermining parasocial relationships. Then the results of the models' prediction comparison are shown in Table II below.

Table. II Comparison of the effects of models in predicting users' parasocial relationships

Data group serial number	Experimental group		Comparison group I		Comparison group II	
	Predictio	Time	Predictio	Time	Predictio	Time
	n	Complexity/s	n	Complexity/s	n	Complexity/s
	Accuracy/%		Accuracy/%		Accuracy/%	
1	87.4	14.26	63.4	243.32	61.3	437.17
2	90.5	14.95	65.8	246.67	57.6	461.48
3	87.9	14.86	65.9	240.64	58.2	476.13
4	92.2	14.77	66.6	226.89	59.4	461.05
5	91.3	14.35	65.3	227.27	61.9	450.36
6	90.1	15.65	63.2	228.86	57.7	460.92
7	89.6	14.34	66.0	239.43	62.1	434.49

8	89.7	15.65	67.1	238.41	61.8	442.36
---	------	-------	------	--------	------	--------

#### IV. DISCUSSION

In order to verify the effectiveness of this model, the above simulation experiments are designed. From the analysis of the data in the above table, it can be seen that the prediction accuracy of the model of the experimental group is higher than that of both comparison groups I and II, and its time complexity is significantly smaller than that of the two comparison groups. The higher time complexity of the model indicates that the processing efficiency of the model is lower. And the prognostic accuracy of the model is lower than 66% after a long time of processing for the two comparison models, which indicates the poor performance of the two models. While the average prediction accuracy of the experimental group model is 89.84% and the average time complexity is 14.85s, the values are better than any one test value of the two comparison group models. In summary, the user parasocial relationship type prediction model based on social media text big data constructed in this paper has the advantages of high prediction accuracy and fast processing speed.

#### V. CONCLUSION

With the rapid development of Internet technology, mobile Internet has also been widely popularized. Social media, as one of the main expressions of the Internet, has become an important part of people's life. Its extensive user participation characteristics have made the research on social media a popular area. Parasocial relationship is considered as a multi-dimensional concept, including friendship, audience-program interaction, and empathy role. In order to improve the accuracy of traditional user parasocial relationship type prediction, this paper constructs a user parasocial relationship type prediction model based on social media text big data. The model performance test results show that the model constructed in this paper has the advantages of high prediction accuracy and accurate prediction of parasocial relationship types.

In the current research, most scholars use social presence to analyze the interaction mechanism between consumers and websites, celebrities and even virtual workers, but there is

little discussion on spatial presence and self-presence. Telepresence is considered to be the degree of psychological and sensory immersion of an individual in a virtual environment. In future studies, presence should be more divided into spatial presence, social presence and self-presence from the perspective of users, so as to further optimize the prediction results of users' parasocial relationship types.

#### References

- [1] Z. Y. Li, H. L. Chen, and X. Huang, "Airbnb or hotel? A comparative study on the sentiment of Airbnb guests in Sydney – Text analysis based on big data," *International Journal of Tourism and Hospitality Management in the Digital Age*, vol. 4, no. 2, pp. 1-10, 2020
- [2] L. Parandeh, F. S. Shafaie, J. Malakouti, M. Mirghafourvand, and M. A. Biostatistics, "The effect of educational text message based on health belief model on osteoporosis preventive behaviors in women: a randomized controlled clinical trial," *Women and Health*, vol. 59, no. 10, pp. 1128-1140, 2019
- [3] U. Wierczyńska-Kaczor, "Exploring the relationship between viewer experience and movie genre – a study based on text mining of online movie reviews," *Problemy Zarządzania - Management Issues*, no.85, pp. 154-175 2019.
- [4] A. Priyadarshi and S. K. Saha, "Web information extraction for finding remedy based on a patient-authored text: a study on homeopathy," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 9, no.1, pp. 8-27, 2020.
- [5] Z. H. Xiao, M. Y. Qiu, and Y. Y. Mei, "Combined prediction model of merchandise sales on the basis of differential evolution algorithm," *Journal of Computational Methods in Sciences and Engineering*, vol. 19, no. 3, pp. 1-11, 2019
- [6] B. B. Wang and Y. X. Ge, "Predicting the influence of Guangfo Metro on the economic level of Foshan City based on the GM (1, 1) model," *IOP Conference Series:*



- Earth and Environmental Science, vol.634, no.1, pp. 012013, 2021.
- [7] J. W. Ma, H. Xu, and H. H. Chen, "Friendship prediction in recommender system," *Journal of National University of Defense Technology*, no. 1, pp. 163-168, 2013.
- [8] S. A. Ray, A. Adisaputera, and I. Pramuniati, "The quality of e-learning based on learning media using Moodle LMS on text of observation reports of grade 10th students of vocational school Telkom Shandy Putra Medan," *Britain International of Linguistics Arts and Education (BIO LAE) Journal*, vol. 2, no.2, pp. 688-699, 2020.
- [9] A. M. Sadiq, H. Ahn, and Y. B. Choi, "Human sentiment and activity recognition in disaster situations using social media images based on deep learning," *Sensors*, vol.20, no. 24, p. 7115, 2020.
- [10] A. K. F. D. Silva, D. G. D. Christofaro, L. M. Vanzella, F. M. Vanderlei, and L. C. M. Vanderlei, "Relationship of the aggregation of cardiovascular risk factors in the parasympathetic modulation of young people with type 1 diabetes," *Medicina (Kaunas, Lithuania)*, vol. 55, no. 9, p. 534, 2019.
- [11] I. Chaplinska, "Functional model of Parasocial relations," *Psychological Journal*, vol. 6, no.5, pp. 66-79, 2020.
- [12] E. L. Cohen, J. G. Myrick, and C. A. Hoffner, "The effects of celebrity silence breakers: liking and parasocial relationship strength interact to predict the social influence of celebrities' sexual harassment allegations," *Mass Communication and Society*, vol. 24, no. 2, pp. 288-313, 2020.
- [13] S. C. Boerman and E. A. V. Reijmersdal, "Disclosing influencer marketing on YouTube to children: the moderating role of para-social relationship," *Frontiers in Psychology*, vol. 10, p. 3042, 2019.
- [14] P. D. Bérail, M. Guillon, and C. Bungener, "The relations between YouTube addiction, social anxiety and parasocial relationships with YouTubers: a moderated-mediation model based on a cognitive-behavioral framework," *Computers in Human Behavior*, vol.99, pp. 190-204, 2019.
- [15] N. Liebers and R. Straub, "Fantastic relationships and where to find them: Fantasy and its impact on romantic parasocial phenomena with media characters," *Poetics*, vol.83, p. 101481, 2020.
- [16] Z. B. Massey, N. C. H. Wong, and J. L. Barbati, "Meeting the (Trans) parent: test of parasocial contact with transgender characters on reducing stigma toward transgender people," *Communication Studies*, vol.72, no. 2, pp. 232, 2021.
- [17] A. M. Yibre and B. Koçer, "Semen quality predictive model using feed forwarded neural network trained by learning-based artificial algae algorithm," *Engineering Science and Technology, an International Journal*, vol. 24, no.2, pp. 310-318, 2020.
- [18] S. Hegde and M. R. Mundada, "Optimized deep neural network based predictive model for customer attrition analysis in the banking sector," *Recent Patents on Engineering*, vol. 14, no.3, pp. 412-421, 2020,
- [19] J. Dahmani, C. Laporte, D. Pereira, P. Bélanger and Y. Petit, "Predictive model for designing soft-tissue mimicking ultrasound phantoms with adjustable elasticity," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 4, pp. 715-726, 2020.
- [20] M. BalaAnand, N. Karthikeyan, and S. Karthik, "Envisioning social media information for big data using big vision schemes in wireless environment," *Wireless Personal Communications*, vol. 109, no.2, pp. 777-796, 2019.
- [21] M. P. S. Chan, K. H. Jamieson, and D. Albarracin, "Prospective associations of regional social media messages with attitudes and actual vaccination: a big data and survey study of the influenza vaccine in the United States," *Vaccine*, vol.38, no. 40, pp. 6236-6247, 2020.

[22] X. H. Xu, X. Yang, X. H. Chen, and B. S. Liu, "Large group two-stage risk emergency decision-making method based on big data analysis of social media," *Journal of Intelligent and Fuzzy Systems*, vol. 36, no. 3, pp. 2645-2659, 2019.

[23] S. Vetova, "Workflow model for big data analysis and visualization", *International Journal of Applied Mathematics, Computational Science and Systems Engineering*, Vol. 3, pp. 99-106, 2021.

Jiatong Meng, masters in applied research to mass media, Graduate School of Humanities, Communications and Social Sciences, Universidad Carlos III de Madrid. Research direction: fan culture, Parasocial interaction of social media users, cross-culture communication.

Yucheng Chen, undergraduate student in Software College of Northeastern University. Expertise: Computer Science and Engineering.

## **Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)**

This article is published under the terms of the Creative Commons Attribution License 4.0

[https://creativecommons.org/licenses/by/4.0/deed.en\\_US](https://creativecommons.org/licenses/by/4.0/deed.en_US)