International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 8942-8953

https,//doi.org/10.70517/ijhsa463766

Network group psychology and youth opinion guidance strategies based on data mining

Minghui Shan[1], Xuezhu Liu[2], Dianqing Jiang[2,*] and Yuqi Qiao[3]
[1] School of Education and Science, Jiamusi University, Jiamusi, Heilongjiang, 123456, China
[2] Library, Jiamusi University, Jiamusi, Heilongjiang, 154007, China
[3] School of Innovation and Entrepreneurship, Jiamusi University, Jiamusi, Heilongjiang, 154007, China
Corresponding authors: (e-mail: tushujiangou@163.com).

Abstract Internet groups refer to netizens who have the same interests and hobbies and hold similar views on an event. In the network group, the youth group occupies the main part, and the formation of the youth public opinion is closely related to the network group psychology. The complexity and diversity of network information often make information mixed with corrupt culture, wrong thoughts, false information and so on. It has certain practical significance for ideological and political education to analyze the characteristics of network group psychology and guide the youth public opinion on this basis. Over the recent years, data mining (DM) technology has developed rapidly. Based on DM technology and support vector machine (SVM) model, this paper conducts targeted research and analysis on network group psychology and youth public opinion guidance, aims to accurately grasp the dynamics of online youth public opinion, and respond promptly, thereby reducing the difficulty of public opinion guidance and improving the efficiency of public opinion guidance. The results show that the public opinion guidance platform based on DM reduces the difficulty of public opinion guidance and improves the guidance efficiency by 6.2%, which has an important reference value.

Index Terms Data Mining, Network Group Psychology, Youth Public Opinion, Guiding Strategy

I. Introduction

Network group psychology is an important part of guiding young people's public opinion. Researching and analysis of network group psychology plays an important role in proposing guidance strategies for youth public opinion. With the advancement of network technology and the complexity of interest, public opinion has an increasing influence on youth groups. It is an important task for scholars from all walks of life to seriously study online public opinion and grasp its occurrence law and social impact. Internet media is also critical in youth public opinion, and DM can use the latest technical means to analyze Internet public opinion, monitor public opinion hotspots in time, predict the potential trends of Internet public opinion, and grasp its new development trends.

With the application in various fields is also increasing, and many scholars have joined the research ranks. Helma C, Cramer T created an OpenAI-gym-like game environment based on DM techniques and proposed a novel A3C+ network for learning RL agents. In experiments, he successfully demonstrated the application of the proposed model in learning 3D fighting games [1]. Xu L, Jiang C studied Privacy Preserving Data Mining (PPDM) in the context of data mining, mainly focusing on ways to mitigate the privacy risks associated with digital mining operations. The results show that PPDM takes a broader perspective on privacy issues and helps to protect sensitive information [2]. Kavakiotis I studied the application of data mining methods in the fields of biotechnology and health sciences, designed a data processing system with diabetes as an example, and finally made accurate prediction and diagnosis of medical data in diabetes [3]. Chaurasia V developed an accurate breast cancer prediction model using data mining techniques. In experiments, he compared the model's classification techniques and prediction results, and the outcomes showed that the model improved the accuracy of breast cancer prediction [4]. Assessing the impact of data mining on fundamental anomalies using a bootstrap approach, Yan X S has found that a number of underlying fundamental signals are substantial predictors of cross-sectional stock returns, even when data mining is taken into account [5]. Slater S, Joksimovic S discussed the tools that data mining has emerged in educational research, and proposed knowledge such as bias tracking tools, text mining, social network analysis, process and sequence mining, which provide useful information for the development of educational models [6]. Ge Z, Song Z reviewed the existing data mining and analysis applications in industry from the perspective of machine learning, and discussed several supervised algorithms for future research on data mining

and analysis in the process industry [7]. The above research on data mining technology is relatively specific, but it does not involve the application of network public opinion guidance.

Analyzing and correctly guiding the Internet public opinion plays an important role in strengthening ideal and political education. Forsman M provides an overall strategy for the construction of public opinion guidance in colleges and universities under the new situation, prevents the fermentation and spread of public opinion in colleges and universities, and creates a harmonious campus cultural atmosphere [8]. Betari R, Junaidi J established a research model combining sentiment analysis and image inpainting. Through the combination of empirical and analysis, he proved the feasibility of the relevant theory and proposed a public opinion guidance strategy for the public subject [9]. Pennavaria K discussed the possibility of population migration in the process of public opinion dissemination and its impact on the starting point of public opinion dissemination through the two processes of "migration" and "transformation" of public opinion dissemination on the Internet. Finally, he discovered the influence of the possibility of public migration and the starting point of public communication in the process of public communication from the perspective of simulation [10]. Given J analyzed online financial public opinion and constructed the online financial public opinion guidance system. Experiments show that financial network public opinion monitoring and specific strategy guidance provide great help to the guidance of financial public opinion [11]. Oh J K studied the network public opinion dissemination events in schools, and by analyzing its basic concepts and status quo, put forward network public opinion guidance strategies and feasible suggestions, and dealt with and resolved the public opinion events in schools [12]. These scholars' research on network public opinion is beneficial to put forward strategies for public opinion guidance, but they do not use DM technology to analyze network group psychology.

In the context of DM, this paper studies and analyzes network group psychology and youth public opinion guidance. In order to make accurate diagnosis and early warning of public opinion texts, it also combines SVM model to classify texts, thereby improving the efficiency of public opinion filtering. In addition, after accurately grasping the dynamics of online public opinion, timely respond to online hotspots and sensitive topics, further improve the ability to respond to and supervise young people's online public opinion, and provide valuable strategies for youth public opinion guidance.

II. Data Mining Technology

DM is an excellent technique for extracting legitimate, new and potentially useful information from the most noisy, incomplete and incomprehensible real-world application data. The extracted information has three characteristics: practical, valid and unknown. The DM process is shown in Figure 1. The types of these data information can be semi-structured, structured, or even heterogeneous.

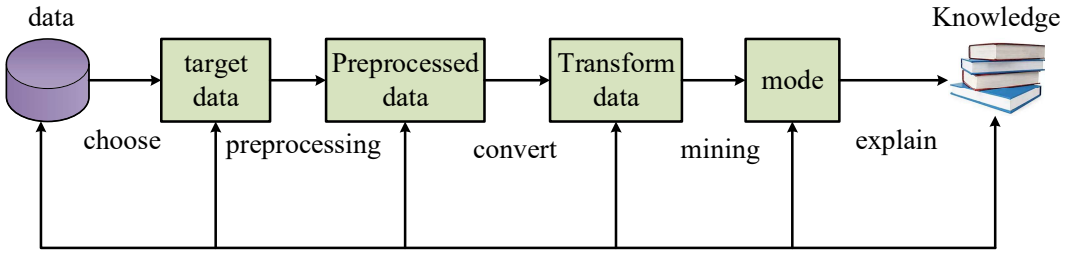

Figure 1: data mining process

Data selection: Determining the operation object of the tracking job, that is, the target object.

Preprocessing: including noise reduction, lost data export, deduplication, data type conversion, etc.

Data Mining: Defining mining tasks such as data summarization, sorting, grouping, correlation detection, or sequential pattern matching, etc., and determine which mining algorithm to use.

Interpretation and evaluation: The algorithms and patterns discovered in the DM stage need to be evaluated by users and machines. Once redundant or irrelevant, they should be removed at this stage so that users can understand and apply them more easily.

III. Network Group Psychology and the Dissemination of Youth Public Opinion

From the perspective of group psychology, the creation and dissemination of online public opinion has the working mechanism of group psychology. Simply understood, the influence of group psychology on network public opinion

can be summarized as the process of forming and disseminating information from multiple sources into a single public opinion. The large amount of information generated in the Internet age brings together groups with common interests or common views, which are called Internet groups [13]. Under the influence of group psychology, diverse, rational and objective information will gradually be replaced by a single polarized emotion, thus forming online public opinion. Figure 2 shows the influence mechanism of network group psychology on network public opinion.

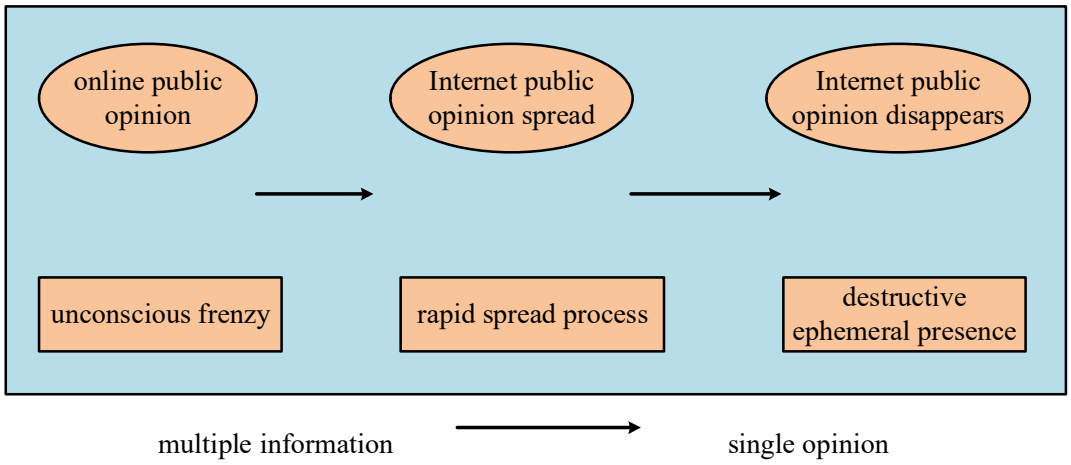

Figure 2: The mechanism of network group psychology on network public opinion

Youth public opinion is the main body of online public opinion dissemination, which originates from uncontrollable fanaticism. Its propagation mode is shown in Figure 3. In the public opinion formation, individuals cannot control public affairs [14]. Online groups are irrational and obtuse, and personal opinions and ideas are eliminated in groups. From a psychological point of view, cognitive impairment and fear of loneliness are the main reasons for herding behavior. Internet public opinion is mostly irrational. Once the online group is under pressure, there will be large-scale chaos, and the way to deal with the chaos is to follow the crowd.

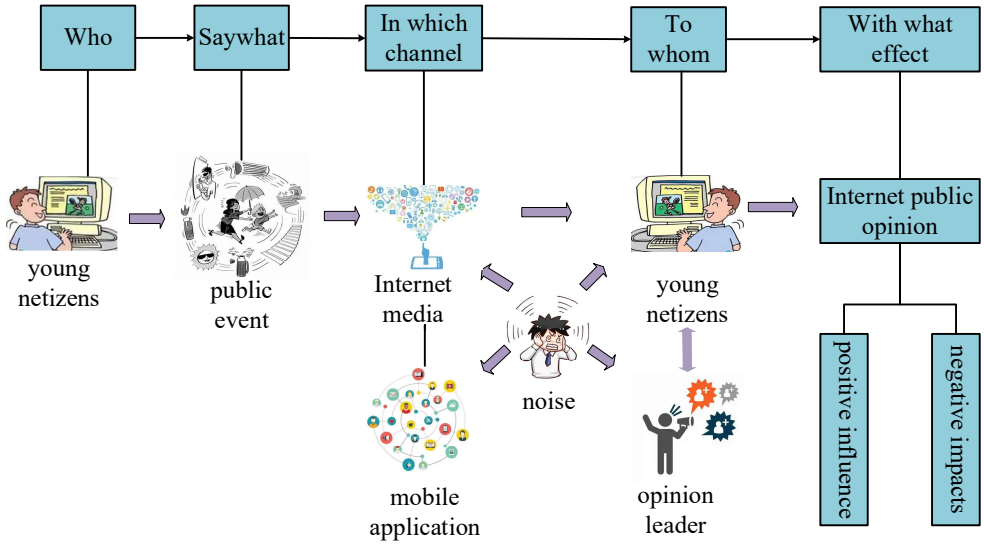

Figure 3: Online youth public opinion dissemination mode

The dissemination of Internet public opinion is an extremely fast and radical information dissemination process. In the process of generating group views, short and lively views are more likely to be accepted. The process of online groups accepting opinions is not an objective and comprehensive acceptance process, but a dissemination process of intercepting and simplifying information [15]. This group tends to accept black-and-white views rather than rationally extracting the truth from a wealth of information. The formation of network groups is composed of a

large number of groups of different natures. Once the network environment changes, the network groups will quickly form new public opinion trends, and group beliefs will also change immediately.

Based on the paranoid and violent psychology of online groups, the spread of online public opinion is a catastrophic short-term process. This group is always full of passion and impulse, always ready for change and destruction. Experiments show that groups can spread responsibility [16]. Once responsibilities are dispersed, restrictions are reduced and the consequences can be unimaginable [17]. The anonymity provided by the Internet makes people more hostile and aggressive than face-to-face communication. In network groups, individuals are more likely to become depersonalized, their self-discipline is weakened, and network group events are more likely to occur.

Most of the youth groups are in school or have just started working. They lack social experience, and are easily over-enthusiastic, often blindly following the trend without thinking. This enthusiasm can indeed play a role in solving the problems of certain events, but it can also be used by some ill-intentioned people to cause some irrational public events and hinder the progress of society. Therefore, it is necessary to detect potential dangers in a timely manner, guide the youth public opinion before it is formed, and occupy the initiative in the guidance of public opinion.

IV. Public Opinion Guidance Platform Based on Data Mining

The data mining youth public opinion guidance platform is an intelligent platform with autonomous processing capabilities. From the simulation model to the intelligent system, the function is more powerful, the degree of intelligence is higher, and the guidance strategy is more scientific. The main functions of the online youth public opinion guidance platform include five aspects: accept public opinion and instructions from the guidance command center; create online comments with specific personality characteristics; work collaboratively with multiple intelligent themes; independently choose public opinion guidance time and guidance methods according to public opinion conditions; specify scientific and reasonable guidance strategies. The public opinion guidance platform based on DM is basically divided into two parts, namely a unified command and control center and multiple distributed public opinion guidance intelligent agents, as shown in Figure 4.

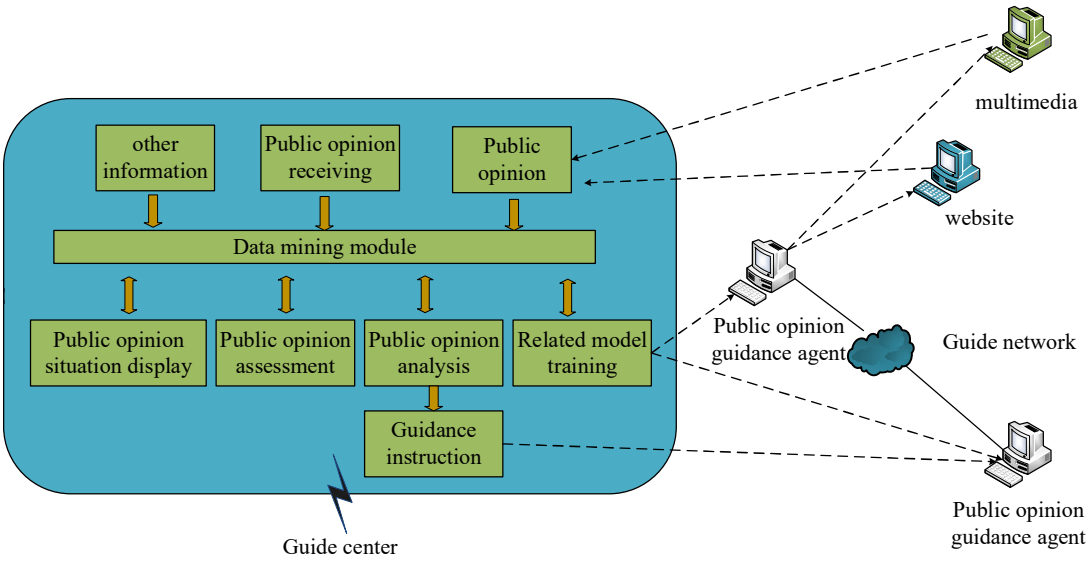

Figure 4: Public opinion guidance platform structure diagram

The Guidance Center is mainly responsible for the collection and aggregation of public opinion information and the comprehensive analysis of public opinion data; on this basis, according to the public opinion analysis index system, the public opinion is analyzed and evaluated, and the visual display is carried out to provide situational information support for the online public opinion struggle; meanwhile, the collected online behavior information and published content information of young users are used to train virtual character models and online comment generation models, and provide them to intelligent agents.

V. Public Opinion Guidance Strategy Model

According to the characteristics of online youth public opinion, the text mining model is adopted in this paper to deeply analyze the online youth public opinion from different levels and angles, so as to propose a more comprehensive and accurate public opinion guidance strategy. The frequently used text mining models is the SVM classification model.

The main concept of an SVM is to create an optimal level of ordering in the vector space where the samples are located. In the vector space, the hyperplane can separate different categories, and can maximize the separation edge of the two types of samples under the premise of satisfying the classification accuracy, and finally use the optimal hyperplane to make decisions on the training sample categories [18]. In classification problems, there are two kinds of samples, one is linearly separable, and the other is linear and inseparable. For linearly inseparable data, the optimal hyperplane cannot be obtained directly from the original vector space. The SVM directly maps the linearly inseparable data to the low-dimensional space, making it linearly separable in the high-dimensional space, and can also obtain the optimal hyperplane in the multi-dimensional feature, and then reorder the data.

(1) Linearly separable problem

Regarding the problem of text categories, the sample is recorded as: $\{(x,y)\}$, among them, y indicates the classification of the sample, and in the classification problem, y has two values. Assuming the training sample set $D=\{(x_1,y_1),i=1,2,.....n\}$, if there is a hyperplane that can distinguish these samples into two types, the plane can be expressed by the following linear formula:

$$\delta_i = y_i(w\cdot x_i + b) \tag{1}$$

$w_i \in w=(1,2,3,4,5...,n)$ is the normal vector that determines the moving direction of the hyperplane, and b is the normal displacement point.

$$\delta_i = \frac{1}{\|w\|}|f(x_i)| \tag{2}$$

If the classification surface (w,b) can accurately separate the two samples on both sides of the hyperplane, that is, for $x_i \in D^n, y_i \in \{+1,-1\}$, if $y_i=+1$, then $w\cdot x_i+b\geq+1$; if $y_i=-1$, then $w\cdot x_i+b\leq-1$, let:

$$\begin{cases} w\cdot x_i+b\geq+1-\zeta_i, y_i=+1 \\ w\cdot x_i+b\leq-1+\zeta_i, y_i=-1 \end{cases} \tag{3}$$

The equal sign holds when the points of the training samples are closest to the hyperplane. As shown in Figure 5:

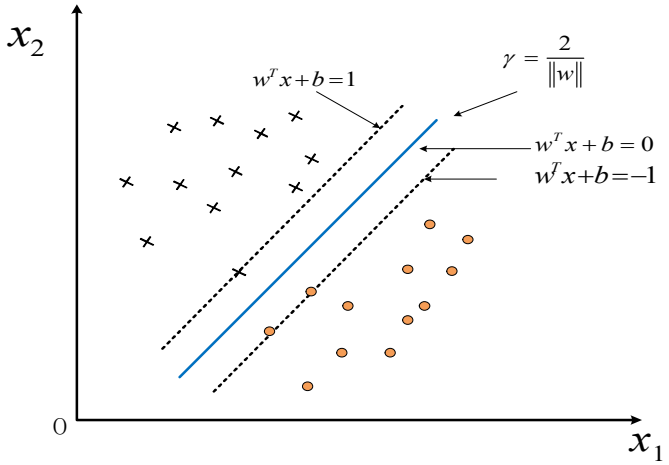

Figure 5: Support Vectors and Intervals

These training sample points in the figure are called "support vectors", and the distance between the support vectors of two different categories to the hyperplane is:

$$\gamma = \frac{2}{\|w\|} \tag{4}$$

This distance is also known as the "separation".

If it wants to find the classification hyperplane with the "maximum interval", it can only make the maximum of ζ, and it needs to find parameters w and b, that is:

$$\begin{aligned} &\max_{\upsilon} \sum_{i=1}^{n} \upsilon_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \upsilon_i \upsilon_j \left(x_i \cdot x_j\right) y_i y_j \\ &s.t. \sum_{i=1}^{n} y_i \upsilon_i = 0 \qquad \upsilon \geq 0, i = 1, 2, ...n \end{aligned} \tag{5}$$

$$\begin{aligned} &\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \zeta_i \\ &s.t. \begin{cases} y_i \left[\left(w \cdot x_i\right) + b\right] - 1 + \zeta_i \geq 0 \\ \zeta_i \geq 0, i = 1, 2, ...n \end{cases} \end{aligned} \tag{6}$$

The result obtained by formula (6) is the model corresponding to the maximum interval hyperplane:

$$f(x) = \operatorname{sgn}\left[\left(w^* \cdot x\right) + b^*\right] = \operatorname{sgn}\left[\sum_{i=1}^{n} \upsilon_i^* y_i \left(x_i \cdot x\right) + b^*\right] \tag{7}$$

$\operatorname{sgn}(\cdot)$ is a symbolic function, w and b are both model parameters. In order to optimize and solve the model, the Lagrange multiplier method [19] was used to establish a function, and the dual problem was obtained, and then solved. The function formula is:

$$J(w, b, \upsilon) = \frac{1}{2} w \cdot w - \sum_{i=1}^{m} \upsilon_i \left\{y_i (w \cdot x_i - b) - 1\right\} \tag{8}$$

Among them, the Lagrange multiplier $\upsilon \geq 0$, $\upsilon = (\upsilon_1, \upsilon_2, ..., \upsilon_n)$, let $J(w, b, \upsilon)$ take the partial derivative of w and b and set it to zero, formulas (9) and (10) can be obtained:

$$w^* = \sum_i \upsilon_i^* y_i b\left(x_i\right) \tag{9}$$

$$b = 1 - w^* \cdot x_i \tag{10}$$

Substituting formula (9) into formula (8) to eliminate w and b in $J(w, b, \upsilon)$. Considering the constraints of formula (10), the dual problem of formula (6) can be obtained by using the duality theorem [20]:

$$\begin{aligned} &\min Q(\upsilon) = \sum_i \upsilon_i - \frac{1}{2} \sum_i \sum_j \upsilon_i \upsilon_j x_i x_j y_i y_j \\ &s.t. \sum_i \sigma_i y_i = 0 \\ &\upsilon \geq 0 \end{aligned} \tag{11}$$

After the model of the optimal classification plane is obtained, the category to which the test sample belongs can be judged according to the optimal hyperplane formula. The relevant formula is as follows:

$$b^* = \frac{1}{N}\left[\sum_{x_i} y_i - \sum_{x_j} y_j \left(x_i, x_j\right)\right] \tag{12}$$

υ_i solved according to formula (11) is the Lagrange multiplier in formula (8), corresponding to the training sample (x_i, y_i). There is an inequality constraint in formula (6), then the above solution process also needs to satisfy the KKT condition, namely:

$$\upsilon_i^* \left[y_i \left(w^* \cdot x_i + b^* \right) - 1 \right] = 0, i = 1, 2 \ldots\ldots n \tag{13}$$

In every training example (x_i, y_i), there are always $\upsilon_i = 0$ or $y_i * f(x_i) = 1$. If $\upsilon_i = 0$, then this sample will not appear in the summation of formula (12); if $\upsilon_i > 0$, then $y_i * f(x_i) = 1$.

(2) Linear inseparable problem

If the collected samples have errors or the samples themselves are inseparable, then the sample set is inseparable whether in the original vector space or in the corresponding high-dimensional space. We consider introducing some parameters to solve this problem so that the SVM can still classify these samples. Assuming there is still an optimal hyperplane that can separate the 2 classes as much as possible, while considering some points that cannot be classified correctly, as shown in Figure 6.

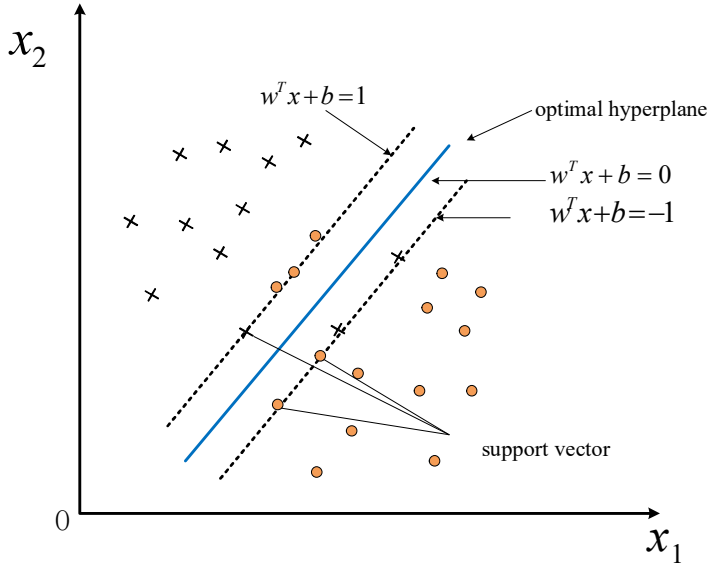

Figure 6: Schematic diagram of data points falling in the wrong area

A loose variable $\zeta_i (\zeta_i \geq 0)$ is introduced for all sample points to measure the distance of the sample from the optimal hyperplane. Then the formula can be written as:

$$y_i (w \cdot x_i + b) - 1 + \zeta_i \geq 0 \tag{14}$$

Furthermore, by entering a non-negative penalty parameter C, the model can be expressed as follows:

$$\begin{aligned} &\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \zeta_i \\ &s.t \begin{cases} y_i \left[(w \cdot x_i) + b \right] - 1 + \zeta_i \geq 0 \\ \zeta_i \geq 0, i = 1, 2, \ldots\ldots\ldots n \end{cases} \end{aligned} \tag{15}$$

Likewise, it first gets its dual problem: for training sample $D = \left\{ (x_1, y_1), i = 1, 2, \ldots\ldots n \right\}$, its dual problem is as follows:

$$J(w,b,\zeta,\upsilon)=\sum_{i=1}^{n}\upsilon_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\upsilon_i\upsilon_j\left(x_i\cdot x_j\right)y_iy_j$$
$$s.t.\begin{cases}\sum_{i=1}^{n}y_i\upsilon_i=0\\ \upsilon_i\geq 0,i=1,2,.....n\end{cases} \tag{16}$$

By finding the optimal solution $\{N_i\}_{i=1}^{n}$ of this equation, the optimal level model when the samples are linearly indistinguishable and errors are allowed can be obtained. Formula (15) shows that the higher the value of the penalty parameter C, the samples will fall into the separation area or the right side, and the higher the penalty for the optimal level, the lower the penalty value C, which means that the more errors are ignored by the classified samples.

(3) Nonlinear mapping when linearly inseparable

According to Cover's theorem, if the samples are not linearly separated from the input region, indirect mapping can be used to place them on a larger size, making the sequentially separated sample points a new high-resolution region. Its specific performance is shown in Figure 7:

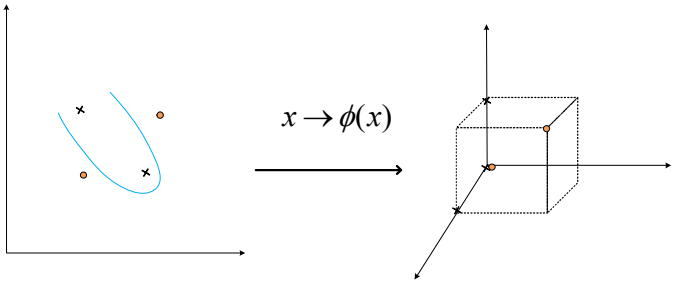

Figure 7: Linearly inseparable and nonlinear mapping

The construction of SVM for text classification can be divided into transforming samples in the original input space into a high-dimensional vector space through nonlinear mapping, and solving the optimal hyperplane in the transformed high-dimensional feature space. The transformation of samples from the original input space to a new high-dimensional feature space requires the help of nonlinear transformations.

Assuming that $\varphi(x)$ represents the attribute vector of x, the model corresponding to the optimal sub-layer can be expressed as:

$$f(x)=\operatorname{sgn}\left[\varphi\left(w^*\cdot x\right)+b^*\right]=\operatorname{sgn}\left[\sum_{i=1}^{n}\upsilon_i^*y_i\varphi\left(x_i\cdot x\right)+b*\right] \tag{17}$$

Among them, w and b are model parameters, then there are:

$$\max_{\upsilon}\sum_{i=1}^{n}\upsilon_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\upsilon_i\upsilon_jy_iy_j\varphi\left(x_i,y_i\right)$$
$$s.t.\begin{cases}\sum_{i=1}^{n}y_i\upsilon_i=0\\ \upsilon\geq 0,i=1,2,.....n\end{cases} \tag{18}$$

It can be seen from the above that the optimal solution of the model can be extended through the core operation of the training mode. Unless the format of the feature map is known, it is impossible to tell what kernel features are appropriate. Therefore, the choice of kernel function has a significant influence on the separation of vector support devices. If the kernel function is not chosen correctly, the samples may be mapped to the wrong range of elements, making the SVM model less useful.

VI. Experimental Test Results of the Public Opinion Guidance Platform

(1) Public opinion evaluation results of the SVM model

Using web crawling technology to scrape youth opinion data from the media, the dataset mainly contains title, content, retweets, and comments. In order to fully display the evaluation effect, three commonly used guiding models, RNN, CNN, and LSTM, are selected for comparison with the SVM model. As shown by the broken line in Figure 8, with the same dataset and the same number of iterations, the SVM model can capture more features of network public opinion in multimedia, while the data sequences of CNN, RNN, and LSTM models have long-term dependencies.

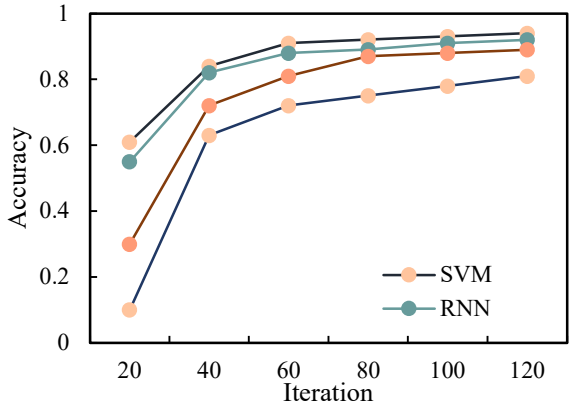

Figure 8: Model accuracy curve

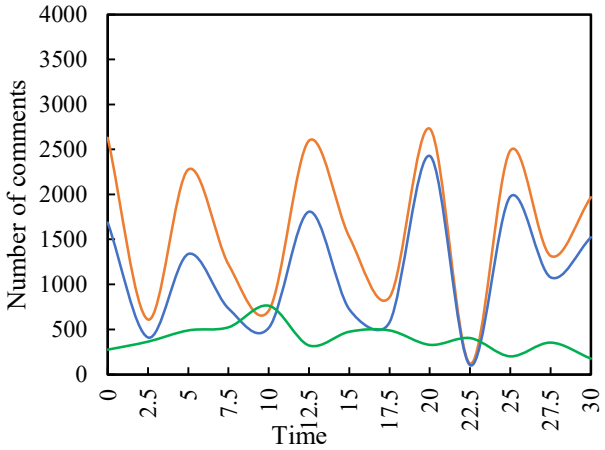

Figure 9: Internet Youth Public Opinion Trends Distribution

The percentages of mildly negative, moderately negative and severely negative public opinion information are shown in Figure 9. An assessment of online youth public opinion and future development status. According to the dataset collected in this paper, the final index of public opinion indicators is used as the input of the SVM model. Actual lines are prediction accuracy curves using the indexing system, which visualize the current status of public opinion events and speculate on future trends. The experimental results indicate that the SVM model improves the accuracy of youth online public opinion prediction and lays a foundation for the formulation of public opinion guidance strategies.

(2) Specificity and sensitivity of the DM public opinion guidance platform to public opinion

The comparison of sensitivity, specificity, and diagnosis rate of the youth public opinion between the DM team and the general team can help effectively grasp the diagnosis ability of the public opinion guidance and dissemination mechanism based on online group psychological events. Figure 10 is a comparison chart of the two groups of data, and Table 1 lists the detailed data.

Table 1: The Particularity and Sensitivity of Youth Public Opinion Guidance

Groups	Specificity (%)	Sensitivity (%)	Diagnosis rate (%)
Data Mining Group	93.4	94.6	99.5
Normal group	87.6	83.7	93.9
t	3.281	3.109	7.665
p	0.008	0.007	0.009

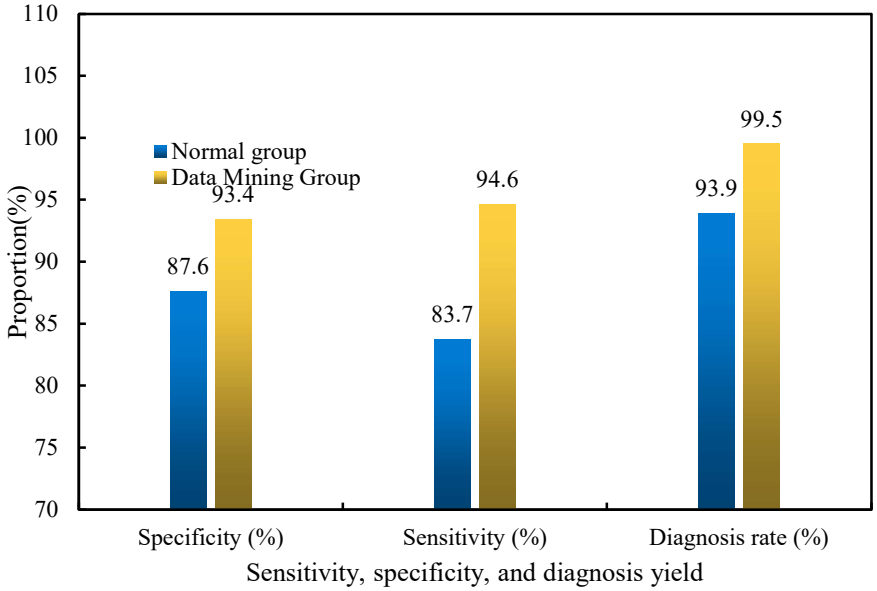

Figure 10: Comparison of sensitivity, specificity, and diagnosis rate of youth public opinion

Figure 10 shows that the DM group's specificity and sensitivity are higher than the normal group. The comparative evaluation results indicate that the DM public opinion guidance platform would enhance the ability of public opinion diagnosis and early warning, and the diagnosis accuracy rate is about 5.9% higher than that of ordinary public opinion guidance.

(3) Guiding difficulty and guiding efficiency of the data mining public opinion guidance platform

The efficiency and difficulty of youth opinion fermentation based on online group psychology can be effectively grasped through the comparison of the difficulty and efficiency of opinion guidance between the DM group and the normal group, as shown in Table 2.

Table 2: Data on the difficulty of opinion guidance and the efficiency of guidance

Groups	Guiding difficulty (R^2)	Guiding efficiency (R^2)
Data Mining Group	0.786	0.937
Normal group	0.889	0.882
t	6.438	6.134
p	0.005	0.005

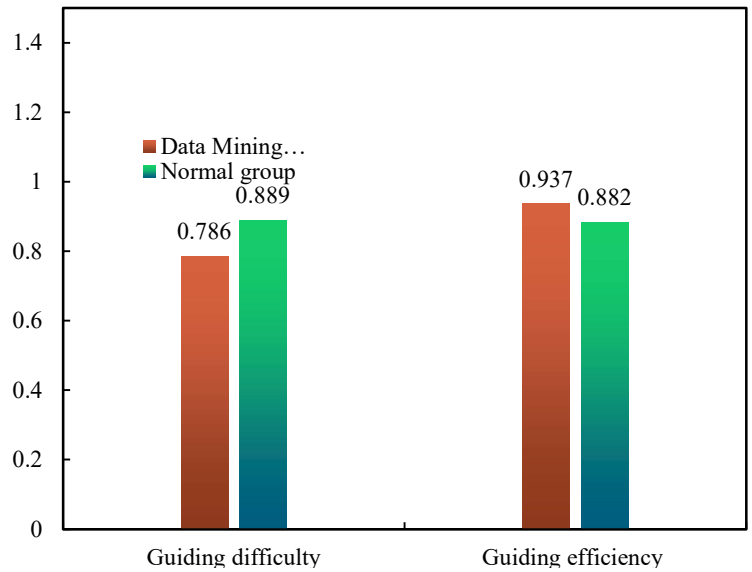

Figure 11: The difficulty and efficiency of public opinion guidance in the two groups

In Figure 11, the DM group is less difficult to guide than the reference group, but more efficient in guiding public opinion than the normal group. The comparison results of public opinion show a significant difference in the efficiency of public opinion guidance between the two groups. The guiding efficiency of the DM group increases by 6.2%, indicating that the public opinion guidance platform and strategy model based on DM can effectively reduce the difficulty of fermenting and guiding public opinion among young and middle-aged people in sudden network groups.

VII. Conclusion

On the basis of DM techniques, in this article studies and analyzes network group psychology and youth public opinion guidance, and combines SVM model to classify and filter public opinion texts. It aims to accurately grasp the current public opinion trends on the Internet, and respond to the hotspots, focuses and sensitive topics of Internet public opinion promptly. In comparison with ordinary opinion guidance mechanisms, the DM-based opinion guidance platform has higher sensitivity, specificity and diagnosis rate for youth public opinion. On the other hand, it reduces the difficulty of public opinion guidance and improves the guidance efficiency, which is helpful to the proposal of youth public opinion guidance strategies in the new era.

Funding

This work was supported 2021 National Library Project for the Arrangement of China Revolutionary Literature and Republican Period Documents, 'Catalogue Summary of Northeast Bookstores 1945-1949', Project Number: Min20210019.

References

[1] Helma C, Cramer T, Kramer S. Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds[J]. J Chem Inf Comput, 2018, 35(4):1402-1411.
[2] Xu L, Jiang C, Wang J. Information Security in Big Data: Privacy and Data Mining[J]. IEEE Access, 2017, 2(2):1149-1176.
[3] Kavakiotis I, Tsave O, Salifoglou A. Machine Learning and Data Mining Methods in Diabetes Research[J]. Computational and Structural Biotechnology Journal, 2017, 15(C):104-116.
[4] Chaurasia V, Pal S. A Novel Approach for Breast Cancer Detection using Data Mining Techniques[J]. Social Science Electronic Publishing, 2017, 3297(1):2320-9801.
[5] Yan X S, Zheng L. Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach[J]. Review of Financial Studies, 2017, 30(4):1382-1423.
[6] Slater S, Joksimovic S, Kovanovi V. Tools for Educational Data Mining[J]. Journal of Educational and Behavioral Statistics, 2017, 42(1):85-106.
[7] Ge Z, Song Z, Ding S X. Data Mining and Analytics in the Process Industry: the Role of Machine Learning[J]. IEEE Access, 2017, 5(2017):20590 - 20616.
[8] Forsman M. Douglas Kellner & Jeff Share (2019). The Critical Media Literacy Guide: Engaging Media and Transforming Education. Leiden: Brill Sense, 125 pp. ISBN: 978-90-04-40452-6[J]. Central European Journal of Communication, 2020, 13(2):299-301.

[9] Betari R, Junaidi J. Upaya Meningkatkan Partisipasi Siswa dalam Pembelajaran Sosiologi Melalui Model Prediction Guide dengan Media Gambar Kelas X IPS 3 SMAN12 Padang[J]. Jurnal Sikola Jurnal Kajian Pendidikan dan Pembelajaran, 2020, 2(1):19-26.
[10] Pennavaria K. INTERNET RESOURCES: Nonprint media preservation: A guide to resources on the Web[J]. College and Research Libraries News, 2019, 64(8):529-532.
[11] Given J. Book Review: The Journalist's Guide to Media Law: A Handbook for Communicators in a Digital World:[J]. Media International Australia, 2019, 173(1):160-160.
[12] Oh J K, Lim H T. A Study on the Media Perception Change toward Tour Guide : Focused on Content Analysis through News Articles[J]. Journal of Tourism and Leisure Research, 2018, 30(8):39-55.
[13] Emoto T, Yamashita T, Kobayashi T. Characterization of gut microbiota profiles in coronary artery disease patients using data mining analysis of terminal restriction fragment length polymorphism: gut microbiota could be a diagnostic marker of coronary artery disease[J]. Heart & Vessels, 2017, 32(1):39-46.
[14] Carneiro N, Figueira G, Costa M. A data mining based system for credit-card fraud detection in e-tail[J]. Decision Support Systems, 2017, 95(MAR.):91-101.
[15] Triguero I, S González, Moyano J M. KEEL 3.0: An Open Source Software for Multi-Stage Analysis in Data Mining[J]. International Journal of Computational Intelligence Systems, 2017, 10(1):1238-1249.
[16] Moustakas, Aristides. Spatio-temporal data mining in ecological and veterinary epidemiology[J]. Stochastic Environmental Research & Risk Assessment, 2017, 31(4):1-6.
[17] Lei, Lin, Qian, et al. Data Mining and Complex Network Algorithms for Traffic Accident Analysis:[J]. Transportation Research Record, 2018, 2460(1):128-136.
[18] Parvizimosaed M, Farmani F, Mons Ef H. A multi-stage Smart Energy Management System under multiple uncertainties: A data mining approach[J]. Renewable Energy, 2017, 102(PT.A):178-189.
[19] Varley J B, Miglio A, Ha V A. High-Throughput Design of Non-oxide p-Type Transparent Conducting Materials: Data Mining, Search Strategy, and Identification of Boron Phosphide[J]. Chemistry of Materials, 2017, 29(6):2568-2573.
[20] Pourghasemi H R, Yousefi S, Kornejady A. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling[J]. Science of The Total Environment, 2017, 609(3):764-775.