Analysis of Media Content Recommendation in the New Media Era Considering Scenario Clustering Algorithm

Lei Tian
Department of Animation Art, Zibo Vocational Institute, Zibo, Shandong, China, 255314
E-mail: 15264327540@163.com

Keywords: scenario clustering algorithm, content recommendation, new media era, communication by microblogging

Received: October 26, 2023

In the era of new media, the abundance of internet information poses a difficulty for users to find media that is both relevant and captivating. Although recommending technologies has made significant progress, it still faces hurdles in dealing with concerns related to data confidentiality and, the algorithmic partiality effect. With the continuous progress of the social economy, new media and micro media are constantly emerging in multiple ways and the methods to access these media contents have become diversified as well. However, it should be noted that diverse types of media content in the era of big data also require excessive time spent in selecting effective content. In response to these demands and defects, a scenario clustering algorithm is introduced in this paper, in which the media content recommendation is taken as the breakthrough point to build a clustering model to express the effective distribution of events by analyzing the network structure and media content distribution model through the analysis of the network structure and the distribution of the media content to represent the effective distribution of events and carry out the comparison of cross-content events, to achieve the effective clustering and analysis of media content. The results of the simulation experiment indicate that the scenario clustering algorithm proposed in this paper is effective and can support the analysis of media content recommendation in multiple dimensions, to provide high-quality media.

Povzetek: Predstavljen je algoritem za gručenje scenarijev za izboljšanje priporočil medijskih vsebin, ob upoštevanju zasebnosti in algoritmične pristranskosti, z analizo mrežne strukture.

1 Introduction

These technologies utilize sophisticated algorithms and analysis of user behavior to create tailored recommendations, hence improving user engagement and satisfaction. Given the abundance of choices available, customers are increasingly looking for personalized experiences. Consequently, the importance of media content recommendations in enabling discovery and applicability becomes crucial. With the continuous progress of the social economy and the popularity of mobile Internet and other devices, people's access to corresponding media content in the new media era has also become diversified, with typical terminals such as cell phones, tablets, computers, and TVs [1-2]. It is attributed to the continuous and in-depth application of big data that media content is no longer released by traditional fixed institutions or people, but common users have become both users and creators of media content [3-4]. How to extract useful information from a huge amount of data information is a highly critical research direction. In the specific new media and micro-media environment, it is a time-consuming and laborious process to distinguish excellent or poor data produced by ordinary users effectively; that is, it is highly challenging to manage and detect the data effectively. In particular, implementing effective measurement of data similarity, cluster analysis, and so on is a prerequisite for ensuring accurate, effective, and healthy media content [5-6]. Efficient media content recommendation algorithms are crucial in assisting users in navigating through a massive amount of information by providing personalized and pertinent suggestions that are customized to their specific tastes, hobbies, and consumption habits. These recommendation methods utilize advanced algorithms, machine learning methods, and analysis of user data to understand user behavior and preferences, resulting in improved user experience and engagement. Industry experts have also carried out a lot of studies on data

management of media content, mainly from effective categorization and clustering of data to implement effective management. Typically, cluster analysis, bag-of-words model, and other ways and means are adopted to achieve effective extraction of data features and apply them to news, online articles, and other clusters. Compared with traditional news, new media content has obvious characteristics: ordinary users create content and two-way transmission. In the above equation ordinary users create content means that most of the content in new media is created by users themselves, and is more rooted in the grassroots, taken from the grassroots, and has a more real-time grasp of topics and trends in public opinion. Through two-way transmission, the users become both the receivers and the creators of new media content. Some studies have suggested that users are attached to their personal preferences, opinions, and relative personal influence when they create new media content [7-8]. This has ensured that new media content in the transmission process allows users can form a certain ship network and become an essential part of the new media time.

However, due to the lowering of the threshold for the creation of media content, the quality of media content gradually varies, and it is extremely important to extract high-quality media content effectively. To address these demands and defects, a scenario clustering algorithm is introduced in this paper. Attempts are made to explore the construction of a recommendation model by sorting out the business logic of media content recommendation in the new media era. Feature analysis and clustering analysis are carried out from the perspectives of network structure and media content. Through the comparison of cross-content events, the effective clustering and analysis of media content is implemented to achieve the effective recommendation of media content.

Methodology	Findings	Limitation
Study [9] proposed an innovative latent genre-aware micro-video recommending algorithm.	The methodology overcomes current methods by taking into account user interests and micro-video content features, leading to recommendations that were more tailored and pertinent.	The computing capabilities necessary for developing and implementing the neural recommendation network might provide practical constraints in real-world scenarios.
Research [10] examined the elements that influence the spread of news material in the digital realm, with a specific emphasis on audience engagement on social media networks including Facebook and Twitter.	The analysis identified specific news concepts and themes that had a higher potential of eliciting audience comments on Facebook and Twitter.	The investigation was confined to examining news content disseminated on Facebook and Twitter, possibly neglecting other social media sites.
Author [11] analyzed the origins of "user-generated content (UGC)" on social media, primarily focusing on strong-tie resources, weak-tie resources, and tourism-tie resources, and their influence on tourist experience within the region.	The findings could imply that UGC sources had an indirect impact on visitor satisfaction by influencing demands, which were then contrasted with visitors' real impressions.	The analysis was dependent on user-generated content sourced from social media, which might introduce bias and fail to consider persons who were not actively involved in social media operations.
Article [12] presented a multi-agent framework that employs an innovative mechanism to rate news articles based on the user's interests obtained from social media.	The method, which they had developed, yields a 28% improvement in outcomes compared to the recommendation systems used by existing news websites.	The efficacy of sentiment analysis in screening news items might fluctuate based on the precision of the assessment algorithm.
Paper [13] investigated the correlation between the user and the API and utilized these associations together with collaborative learning approaches	The experiments were conducted using a dataset from the actual world. The findings of the experiments demonstrate that their models outperform all other	The effectiveness of the models was highly dependent on the quality and expressive capability of the vector depictions provided by doc2vec. However, it was

to provide API recommendations.	methods that were evaluated.	possible that these representations might not fully capture all the subtleties of user-API interactions.

1.1 Data collection

We gathered the dataset from kaggle https://www.kaggle.com/datasets/pypiahmad/social-recommendation-data. The Social Recommendation Database consists of ratings combined with social or trusted connections among users from two distinct networks - LibraryThing (a website for book reviews) and Epinions (a platform for general consumer reviews). The collection combines review data with social relationships between users, offering a distinct chance to examine the impact of networking sites on rating patterns and vice versa. Table 1 shows the dataset description.

Table 1: Dataset description

	Library Thing	Epinions
Number of users:	73,882	116,260
Number of items:	337,561	41,269
Number of ratings/feedback:	979,053	181,394
Number of social relations:	120,536	181,304

2. Data preprocessing

2.1 Data cleaning

The data cleansing procedure for data manipulation encompasses the elimination of duplicates, management of missing values, and maintenance of data integrity across various platforms. More precisely, it involves removing duplicate entries, filling in missing grades or user connections when feasible, and standardizing user identification. Furthermore, outlier detection can be utilized to eliminate atypical ratings or associations in data.

Z-score normalization

The approach described is the most widely used normalization technique, which transforms all input values into a normalized statistic with a mean of zero and a standard deviation of one.

For every attribute, the mean and standard deviation are computed. The normalization process involves utilizing the calculated mean and standard deviation to standardize each value of property$w$. The equation

representing the transition is provided as

$$z = \frac{(w - mean\,(w))}{std}(w) \quad (1)$$

The notation mean (w) represents the average value of the attribute$w$, while $std(w)$ represents the measure of dispersion known as the standard deviation of the characteristic$w$. The benefit of the approach emerges from its ability to mitigate the impact of abnormalities on the data.

2.2 Scenario clustering algorithm

The so-called scenario clustering algorithm is designed in response to the new media environment to implement the conversion of media content data and contact data, support the clustering analysis of content and structure from the complex network, and perform the similarity calculation process. A Scenario Clustering Algorithm classifies data into clusters according to similarities in predefined situations. The process generally includes procedures such as extracting features, measuring similarity, and utilizing clustering techniques including k-means or hierarchical clustering. The system assesses patterns and correlations among scenarios to categorize them properly, facilitating comprehension of data structures and enabling predictions. By employing iterative refinement, it improves both accuracy and efficiency, making it applicable in many domains including data processing, pattern recognition, and decision-making. The versatility of this tool enables it to be easily adapted to a wide range of datasets, hence facilitating problem-solving and the finding of insights in difficult settings. Its specific process is shown in Figure 1 as the following.

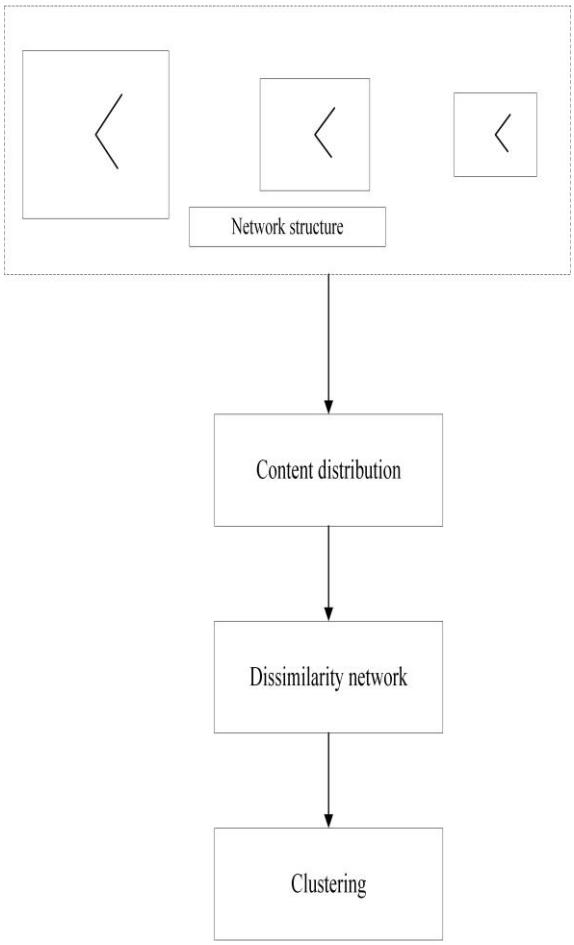

Figure 1: Network event similarity model based on the network structure entropy and the content distribution entropy.

Starting with the data content, the similarity of media content can be measured in two aspects as the following: one aspect is the structural entropy of the network, and the other aspect is the distributional entropy of the content. The structure of the network is used to measure the similarity of the topological chain formed by the communication in the new media era, and the partial entropy of the content mainly measures the similarity of the content changes during the communication process. The similarity measure based on the network structure can be calculated and evaluated quantitatively from the complexity and similarity of the network during the propagation process, and the similarity measure of the content is mainly conducted quantitatively based on a fixed model.

2.3 Similarity measure based on entropy

In the new media era, during the media content dissemination under the guidance of important nodes, ordinary nodes with time gradually present the scale of decreasing. The details are shown in Figure 2 as the following.

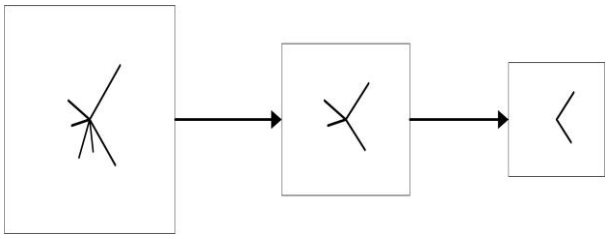

Figure 2: Evolution of the communication of new media events.

(1) Similarity based on the entropy of network structure

In essence, the measurement of network structure entropy is the calculation and analysis of probability using nodes. In this paper, divergence is introduced to conduct the specific quantitative analysis, which is calculated according to equation (1) as the following.

$$D_{KL}\left(p\|q\right)=\sum_{i=1}^{N}p\left(x_i\right)*\log\frac{p\left(x_i\right)}{q\left(x_i\right)} \qquad (2)$$

In the above equation, the probabilities of the distribution of specific dimensions are denoted by p and q.

Kullback-Leibler (KL) divergence, also known as relative entropy, is a measure of the similarity of two distributions [14-15].

After the similarity of network size and network topology in the transmission process are taken into comprehensive consideration, the specific quantitative calculation is carried out, as shown in equation (2) below.

$$D_n\left(g_1,g_2\right)=1-w_1\sqrt{\frac{EMD\left(\mu_{g1},\mu_{g2}\right)}{\log 2}}+w_2\left|\sqrt{NND\left(g_1\right)}-\sqrt{NND\left(g_2\right)}\right| \qquad (3)$$

The specific calculation of the discrete point index for the MVD network is shown in equation (3) as the following.

$$NND(G)=\frac{J(P_1,P_2,\cdots,P_N)}{\log(d+1)} \tag{4}$$

Normalized conversion is carried out based on equation (3), and the details are shown in equation (4) as the following.

$$J(P_1,P_2,\cdots,P_N)=\frac{1}{N}\sum_{i,j}p_i(j)*\log\left(\frac{p_i(j)}{\mu_j}\right) \tag{5}$$

The specific calculation μ_j is shown in equation (5) as the following.

$$\mu_j=\sum_{i=1}^{N}p_i(j)/N \tag{6}$$

(2) Similarity based on the entropy of content distribution

Similarity measurement by using specific structural entropy is a measure of event similarity from the topology of the event, where specific individual node probability distributions instead of the entire transmission network are used and corrected based on the practical scale of media content.

The specific calculation of the network similarity model based on entropy is shown in equation (6) as follows.

$$D(g_1,g_2)=w_1*D_{l(g_1,g_2)}+w_2*D_{n(g_1,g_2)} \tag{7}$$

3. Event clustering model based on NRL and K-means

From the similarity calculation model proposed in this paper, the distance between events or the distance matrix between multiple events can be derived directly. The specific calculation of the basic attributes is shown in equation (7) as the following.

$$E_i=\{M_i,I_i\} \tag{8}$$

In the above equation, the network event is denoted by specific, and the numbers of media are denoted by M_i and I_i, respectively, and all the above parameters need to be normalized.

3.1 Improved K-means algorithm

The K-means algorithm requires that the number of clustering centers should be obtained in advance. However, in most cases, it is not possible to know the exact number before clustering. If the number of clustering centers is taken unreasonably, it can increase the error of clustering results [16-17]. Before clustering, a coarse clustering of historical run scenarios based on the Canopy algorithm is required to determine the number of clustering centers. The Canopy algorithm does not require specifying the number of clusters in advance, and coarse clustering of the data can generally be conducted in the preprocessing stage. The Canopy algorithm can be used to optimize the clustering results by accurately processing the data according to the coarse clustering results. The specific steps of the Canopy algorithm are described as the following:

Step 1: Input the set List composed of original data, and set the distance threshold T1 and T2, and T1>T2.

Step 2: Select the data point P' from the List randomly, take point P' as the first data center Canopy, and remove it from the List.

Step 3: Take a point Q from the List and calculate the distance from the point Q to all the Canopy that has been generated. If the distance from point Q to a Canopy is less than T2, add point Q to that Canopy and remove it from the List; that is, point Q is considered to be close enough to that Canopy to be the center of other Canopies. If the distance of point Q to all Canopy is greater than T1, then point Q is added as a new Canopy and removed from the List. If the distance from point Q to some Canopy is between T2 and T1, point Q is added to that Canopy. However, it will not be removed from the List and will continue to be included in the subsequent calculations.

Step 4: Repeat Step 3 for the other points in the List until the List is empty.

The number of coarse clusters obtained from the output of the Canopy algorithm is taken as the input parameter of K-means clustering to obtain the final clustering result.

3.2 Generation of typical scenarios

Due to the uncertainty of the scenario, it is necessary to cluster the historical output curves of the new energy power sources in a year to obtain the typical output curves. However, based on the improved clustering method that analyzes each historical output curve separately, it may result in a different number of clustering curves for each power source, which can increase the complexity and computational effort of the subsequent analysis. Hence, it is necessary to define power supply operation scenarios, that is, the scenarios of power output characteristics obtained by taking all new energy sources in the system as a whole. In the clustering process of each curve, the first Canopy coarse clustering is carried out for each power supply to obtain the optimal number of clusters; then K-means clustering is performed for the power supply operation scenario to get the typical operation scenario of the new energy power supply. The specific process is described as the following.

Firstly, the historical curves of n new energy power sources are analyzed by Canopy coarse clustering, and the number $k_i\left(1 \le i \le n'\right)$ of coarse clustering centers of each power source is obtained accordingly. The number of coarse clusters with the most occurrences among all the coarse clusters is calculated, and this number of cluster centers is used as the optimal cluster number (k) for the typical operation scenario of the power supply. Subsequently, k is used as the input parameter of the next K-means clustering method to perform uniform scenario clustering of power supply operation scenarios and obtain typical operation scenarios of power supplies. At the same time, the occurrence probability (P) of each typical scenario can be obtained based on the number of historical scenarios contained in each typical scenario. As the clustering process of typical scenarios of loads is similar to that of the power supplies, it will not be described in detail herein[18-19].

After the clustering results of power (it is assumed that there are m typical scenarios) and load (it is assumed that there are n_0 typical scenarios) are obtained separately, the typical pairwise combination of scenarios of power and load is carried out, and the probability of occurrence of $m \times n_0$ typical system operation and each system operation scenario $P_i \times P_j\left(1 \le i \le m, 1 \le j \le n_0\right)$ can be obtained. The process of system typical operation scenario generation is shown in Figure 3 as the following. The scenario has covered the operation scenarios of the possible power sources containing the source-load matching scenarios of the power output characteristics of each new energy source.

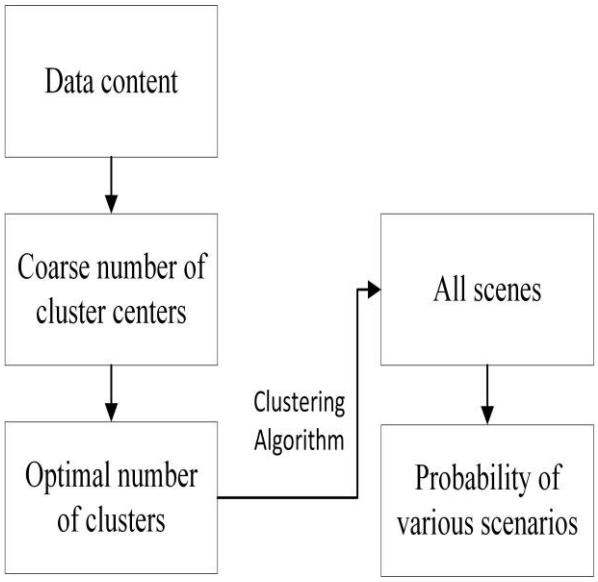

Figure 3: Flow chart for the generation of typical scenarios.

Partitioning is carried out on a power system, and flexibility assessment is conducted within and between the regions of the power system, respectively. The assessment indexes are shown in Figure 4 as the following. The intra-regional flexibility assessment index includes the partitioned supply and demand upward and downward flexibility deficiency index and the partitioned grid flexibility deficiency index, and the inter-regional flexibility assessment index includes the partitioned transmission channel flexibility deficiency index.

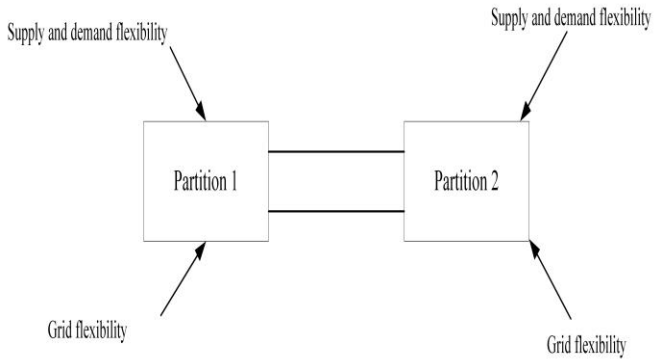

Figure 4: Schematic diagram of flexibility assessment indexes.

3.3 Flexibility indexes for upward and downward adjustment of supply and demand by region

The flexibility index for the supply and demand by region is an index to determine whether the flexibility resources in the region are all meeting the flexibility demand. The schematic diagram of the flexibility resources in the region is shown in Figure 5 below, in which wind power, PV, and hydropower are uncontrollable units, and thermal power and energy storage are controllable units. In the calculation of flexibility demand, the fluctuation of the output of uncontrollable units and load is taken into consideration. In the calculation of flexibility supply, the flexibility supply capacity provided by controllable units is taken into consideration [20-21].

The specific calculation of the power demand and supply generated by the uncontrollable units and loads in the partition at moment t is shown in equation (8) and equation (9) as the following.

$$P_{uncon\_demand}(t) = P_{load}(t) + \Delta P(t) \quad (9)$$

$$P_{uncon\_supply}(t) = P_{wind}(t) + P_{PV}(t) + P_{hydro}(t) \quad (10)$$

When $\Delta P(t) > 0$ it is considered that the partition delivers power to the outside world and increases the power demand; $\Delta P(t) < 0$ it is considered that the partition receives power from the outside world and decreases the power demand.

The upper and lower limits of the power demand variation in the uncontrollable part are calculated. The details are shown in Equation (10) and Equation (11) as the following.

$$P_{uncon\_demand\_max}(t) = (1+\lambda)P_{uncon\_demand}(t) - (1-\lambda)P_{uncon\_supply}(t) \quad (11)$$

$$P_{uncon\_demand\_min}(t) = (1-\lambda)P_{uncon\_demand}(t) - (1+\lambda)P_{uncon\_supply}(t) \quad (12)$$

In the above equation: λ stands for the power fluctuation coefficient; the larger the λ is, the greater the power fluctuation is.

Based on the variation range of power demand described above, the flexibility demand calculation process of the partition is as the following.

When $P_{uncon\_demand\_min}(t) > P_{uncon\_demand}(t-1)$ only upward adjustment of flexibility demand is calculated as shown in equation (12) as the following.

$$P_{demand\_up}(t) = P_{uncon\_demand\_max}(t) - P_{uncon\_demand}(t-1) \quad (13)$$

When $P_{uncon\_demand\_max}(t) > P_{uncon\_demand}(t-1) > P_{uncon\_demand\_min}(t)$, there are both upward and downward flexibility demands, which are calculated according to equation (13) as the following.

$$\left.\begin{aligned} P_{demand\_up}(t) &= P_{uncon\_demand\_max}(t) - P_{uncon\_demand}(t-1) \\ P_{demand\_down}(t) &= P_{uncon\_demand}(t-1) - P_{uncon\_demand\_min}(t) \end{aligned}\right\} \quad (14)$$

When $P_{uncon\_demand}(t-1) > P_{uncon\_demand\_max}(t)$, it is necessary to adjust flexibility demand downward, which is calculated according to equation (14) as the following.

$$P_{demand\_down}(t) = P_{uncon\_demand}(t-1) - P_{uncon\_demand\_min}(t) \quad (15)$$

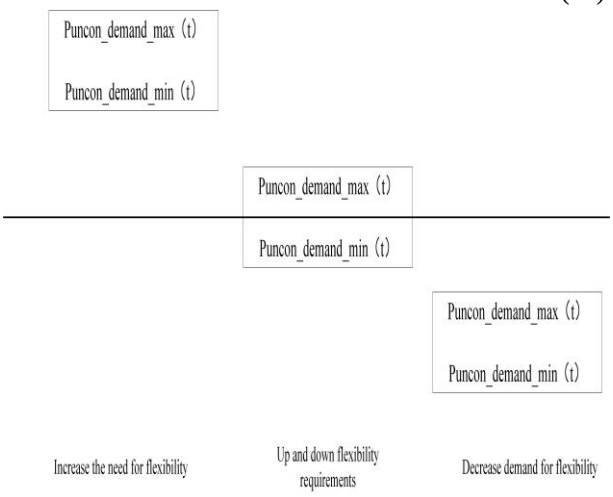

Figure 5: Diagram of partition flexibility resources.

The upward and downward flexibility supply by region is provided by the controllable unit, and it can be calculated according to Equation (15) and Equation (16) as follows.

$$P_{\text{supply\_up}}(t) = P_{con\_gen\_\max} - P_{con\_gen}(t) \quad (16)$$

$$P_{\text{supply\_down}}(t) = P_{con\_gen}(t) - P_{con\_gen\_\min} \quad (17)$$

In the above equation: $P_{con\_gen}(t)$ stands for the output value of controllable units; $P_{con\_gen\_\max}$ and $P_{con\_gen\_\min}$ stands for the maximum output value and minimum output value of all controllable units, respectively.

Based on the upward and downward flexibility for demand and supply described above, the upward and downward poor flexibility indexes can be obtained and calculated according to equation (17) and equation (18) as follows:

$$F_{up}(t) = \frac{P_{demand\_up}(t)}{P_{\text{supply\_up}}(t)} \quad (18)$$

$$F_{down}(t) = \frac{P_{demand\_\text{down}}(t)}{P_{\text{supply\_down}}(t)} \quad (19)$$

The upward and downward poor flexibility indexes indicate the capacity of the flexibility resources to meet the upward and downward flexibility demands. When $F_{up}(t) < 1$, $P_{\text{supply\_up}}(t) > P_{demand\_up}(t)$, and the flexibility resources have a certain margin. When $F_{up}(t) > 1$, $P_{\text{supply\_up}}(t) < P_{demand\_up}(t)$, and the flexibility resources may not be able to meet the demand of the grid. Hence, it is necessary to take measures such as allocating controllable units, energy storage, new energy, or load removal to keep the balance of supply and demand of the flexible resources. Similarly, when $P_{down}(t) > 1$, the flexibility resources fail to meet the demand of the grid, it is necessary to take measures such as reducing new energy and increasing load to keep the supply and demand of flexibility resources in balance [22-23].

3.4 Poor flexibility index of partition grids

Access to a large number of new energy sources can affect the tidal distribution of the system, and the partition grid flexibility index is an indicator that can be used to determine whether the grid structure and line transmission capacity in the region can meet the tidal distribution. Its weighted average of the N branches with the largest calculated load factor in the network at moment t[24-25]. Thus, the poor grid flexibility index at moment t can be calculated according to equation (19) as the following.

$$Flex\_net(t) = \sum_{i=1}^{N} \mu_i L_i(t) \quad (20)$$

In the above equation: i stands for an arbitrary branch; μ_i stands for the flexibility weighting factor; $L_i(t)$ and stands for the calculated load factor. The detailed calculation is shown in Equation (20) and Equation (21) as the following:

$$\mu_i = \frac{\sigma_i^2}{\sum_{i=1}^{N}\sigma_i^2} = \frac{\sum_{t=1}^{T}\left(L_i(t) - \overline{L}_i\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(L_i(t) - \overline{L}_i\right)^2} \quad (21)$$

$$L_i(t) = \frac{S_i(t)}{S_{i\max}} \quad (22)$$

In the above equation: σ_i^2 stands for the variance of the calculated load factor; $\overline{L}_i$ stands for the average of the calculated load factor overall moments; T stands for the maximum number of moments; $S_i(t)$ stands for the transmission capacity; and $S_{i\max}$ stands for the maximum transmission capacity. It is possible to identify the "defects" that restrict the flexibility of the network frame by comparing the branch circuits horizontally to resist the fluctuation of flexibility resources based on μ_i. When $L_i(t) > 1$, overload can

occur in the actual operation, and measures such as wind prevention, light prevention, and load cutting should be taken. When $L_i(t)<1$, the branch circuit can be operated normally.

According to the definition, $Flex\_net(t)>0$. When $Flex\_net(t)>1$, one or more branches can be overloaded in actual operation it is necessary to take appropriate measures to cope with the situation. Hence, the smaller the Flaa_net(t) is, the better the flexibility of the grid is.

3.5 Poor flexibility index for partition transmission channel

Some regions are connected with more new energy sources and less load, and the demand in the region cannot be generated from energy sources, and a large volume of electric power needs to be transmitted to other regions. In this regard, the flexibility index of the transmission channel is defined to determine whether the transmission capacity of the transmission channel meets the outgoing power.

By the power flow relationship between the partition, the transmission from the two ends of the partition is divided into the sending end and the receiving end [26-27]. It is assumed that there are n transmission lines between the sending end and the receiving end, and define the active power of lines 1-n as $P_1, P_2, \cdots, P_n$, respectively, which constitute the transmission channel, as shown in Figure 6 below.

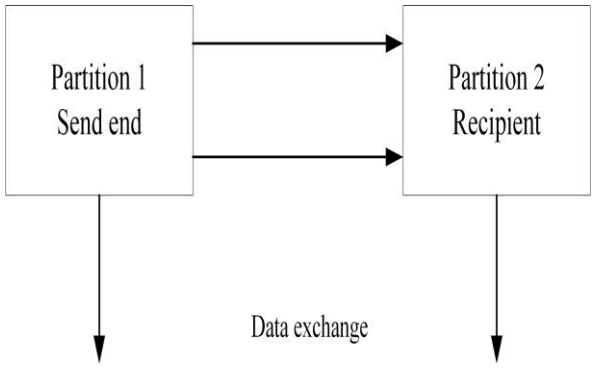

Figure 6: Schematic diagram of the power transmission channel.

About the sending end, the specific calculation of the sending power at moment t is shown in equation (22) as the following:

$$P_{supply}(t)=P_{wind}(t)+P_{PV}(t)+P_{hydro}(t)+P_{con\_gen}(t)-P_{load}(t)-\Delta P(t) \tag{23}$$

(1) Normal operating conditions

In the case where the transmission loss is ignored, the delivered power at the sending end is the transmission power of the inter-regional transmission channel, and the specific calculation is shown in equation (23) as the following.

$$P_{\text{supply}}(t)=P_{trans}(t)=P_1(t)+P_2(t)+\cdots+P_n(t) \tag{24}$$

The power distribution coefficients of lines 1 to n in the transmission channel are defined as $r_1, r_2, \cdots, r_n$, respectively, and their specific calculation is shown in equation (24) as the following:

$$P_{trans}(t)=P_1(t)+P_2(t)+\cdots+P_n(t)=$$
$$r_1P_{trans}(t)+r_2P_{trans}(t)+\cdots+r_nP_{trans}(t)=\sum_{i=1}^{n}r_iP_{trans}(t) \tag{25}$$

In the above equation: r_i stands for the power distribution coefficient of the ith transmission line, and

$$\sum_{i=1}^{n}r_i=1.$$

In different operation scenarios of the system, the power distribution of the transmission line is different. However, in the same operation scenario, the power distribution coefficient of the transmission line is considered to be the same. The definition r_i and its specific calculation are shown in equation (25) as the following.

$$r_i=\frac{\overline{P}_i}{\sum_{i=1}^{n}\overline{P}_i} \tag{26}$$

The above equation: $\overline{P}_i$ stands for the average power of the ith transmission line at all moments, and the detailed calculation is shown in equation (26) below.

$$\overline{P}_i = \frac{1}{T}\sum_{t=1}^{T} P_i(t) \tag{27}$$

Among all transmission lines in the transmission corridor, if one transmission line i reaches the upper power limit for transmission, the specific calculation is shown in equation (27) as the following.

$$r_i P_{trans}(t) = P_{i\max} \tag{28}$$

In this case, the total transmission power between the sending end and the receiving end is considered to have reached its maximum value. The value of the transmission capacity of the transmission channel when each line reaches the upper limit of power is calculated, and the minimum value is taken as the upper limit of transmission capacity of the transmission channel. The specific calculation is shown in equation (28) as the following:

$$P_{normal\_\max} = \min_{1<i<n} \frac{P_{i\max}}{r_i} \tag{29}$$

The specific calculation of the maximum output power at the defined delivery end is shown in equation (29) as the following.

$$P_{source\_\max} = \max\left(t \middle| P_{wind}(t) + P_{PV}(t) + P_{hydro}(t)\right) + \min\left(t \middle| P_{con\_gen}(t)\right) - \min\left(t \middle| P_{load}(t)\right) - \min\left(\Delta P'(t)\right) \tag{30}$$

In the above equation, $\Delta P'(t)$ stands for the exchange power between the sending end and other partitions (excluding the receiving end).

Under normal operation conditions, the specific calculation of the transmission channel flexibility inflexibility index is shown in equation (30) as follows.

$$F_{channel} = \frac{P_{source\_\max}}{P_{normal\_\max}} \tag{31}$$

If $F_{channel} > 1$, it means that the transmission channel is not flexible enough to meet the transmission demand of the maximum output of the power supply at the sending end, new transmission lines are required. If $F_{channel} = 1$, it means that the transmission channel just meets the transmission demand of the maximum power output at the sending end. If $F_{channel} > 1$, it means that the transmission channel is flexible enough and the power output at the sending end is not blocked.

(2) n-1 operating condition

If the k-th line in the transmission channel fails and is withdrawn from operation, only n-1 transmission lines are operating between the sending end and the receiving end. The specific calculation is shown in equation (31) as the following:

$$P_{trans}(t) = P_1(t) + P_2(t) + \cdots + P_n(t) - P_k(t) = r_1 P_{trans}(t) + r_2 P_{trans}(t) + \cdots + r_n P_{trans}(t) - r_k P_{trans}(t) = \sum_{i=1, i\neq k}^{n} r_i P_{trans}(t) \tag{32}$$

If one transmission line i in the n-1 transmission lines in operation reaches the upper power limit, the total transmission power between the sending and receiving ends is considered to have reached the maximum value. The value of the transmission capacity of the transmission channel when each line reaches the upper limit of power is calculated, and the minimum value is taken as the upper limit of the transmission capacity of the transmission channel [28]. The specific calculation is shown in equation (32) as the following:

$$P_{k\_fault\_\max} = \min_{1<i<n, i\neq k} \frac{P_{i\max}}{r_i} \tag{33}$$

In the above equation: $P_{k\_fault\_\max}$ stands for the maximum value of the power transmitted by the transmission channel when the k-th line fails.

In each operation scenario, the minimum value of the power that can be transmitted by the remaining transmission lines in case of failure of each transmission line in the transmission channel is calculated as the upper limit of the power transmitted by the transmission channel under n-1 operation conditions between these two partitions. The specific calculation is shown in equation (33) as the following:

$$P_{fault\_\max} = \min_{1<k<n} P_{k\_fault\_\max} \tag{34}$$

Under n-1 operating conditions, the transmission channel's poor flexibility index can be calculated. The specific calculation is shown in equation (34) as the following:

$$P_{\text{channel}\_fault} = \frac{P_{source\_\max}}{P_{fault\_\max}} \tag{35}$$

The meaning of P_{channel_fault} is similar to that of P_{channel}, except that P_{channel_fault} applies to n-1 operating conditions, whereas P_{channel} applies to normal operating conditions.

3.6 Flexibility Assessment Process

Based on the improved K-means algorithm to generate several types of typical scenarios, the flexibility evaluation results of each type of scenario are analyzed, and the comprehensive flexibility evaluation index is weighted based on the probability of occurrence of each type of scenario [29-30]. The specific process is shown in Figure 7 as the following.

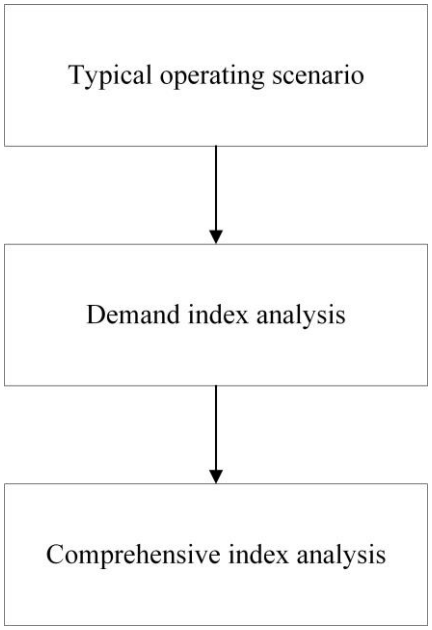

Figure 7: Flow chart of flexibility assessment.

4. Example analysis of the model

4.1 Overview of the Events

Based on the demand of the model for the relevant data, the average number of media, the average influence, and the average amount of original content in each blog post within the event are calculated, and descriptive statistics are carried out. The results are shown in Figure 8 and Figure 9 as the following.

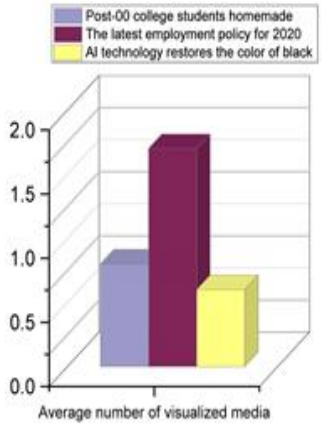

Figure 8: Overview of the basic attributes of events (partial).

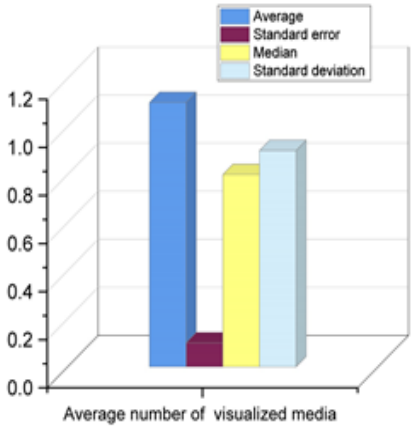

Figure 9: Descriptive statistics of the basic event attributes.

From the results, it can be observed that the media content is stable at around 1 tweet (microblogging message), with a relatively great gap generated depending on the event or topic.

4.2 Similarity Measure Results Based on Entropy

The network structure entropy and event content entropy (that is, NND measurement index) of each event in the dataset are shown in Figure 10 as the following.

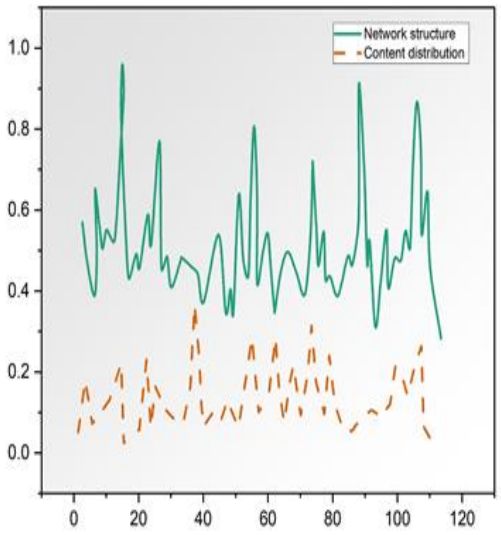

Figure 10: Entropy of the event network structure versus the entropy of the content distribution.

It can be observed from the results in Figure 10 above that the structural EMD and the text distribution (content distribution) EMD are consistent in general. However, there are relatively significant differences in individual events.

5. Results of event clustering

5.1 Clustering results in the experimental group

The clustering results are plotted based on the SSE (Sum of the Squared Errors) to obtain the optimal number of clustering categories. The details are shown in Figure 11 as the following.

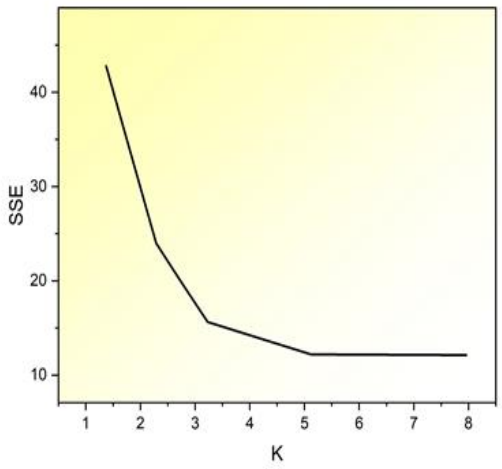

Figure 11: Number of clustering categories of the SSE index in the experimental group.

When k=4, the SSE index is decreased rapidly. At this point, k corresponds to a more realistic number of clustering categories at this time.

Through the observation of the raw data, the typical events and characteristics of each category in the final clustering results can be obtained, as shown in Figure 12 as the following.

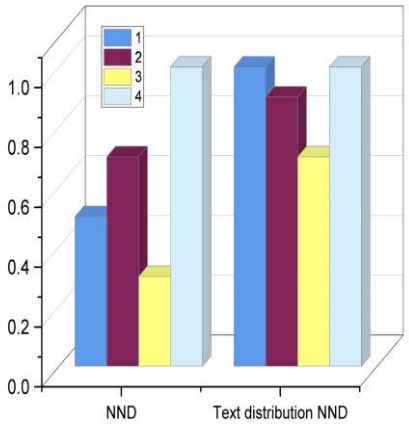

Figure 12: Description of the clustering categories in the experimental group.

The NND values in the table are the normalized values of the mean NND values in this class. It can be observed from the results that the propagation structure of event 1 is relatively homogeneous, and the local text cannot characterize the whole well. Event 2 includes a large number of events and relatively complex dissemination. Event 3 has local information that can better characterize the whole. Event 4 has local information that cannot well characterize the whole. The events have triggered many controversies and discussions, and the dissemination network-N BH structure is irregular.

5.2 Clustering results in the control group

In the control group, the same SSE index is used to identify the optimal number of clustering categories, and the results are shown in Figure 13 as follows.

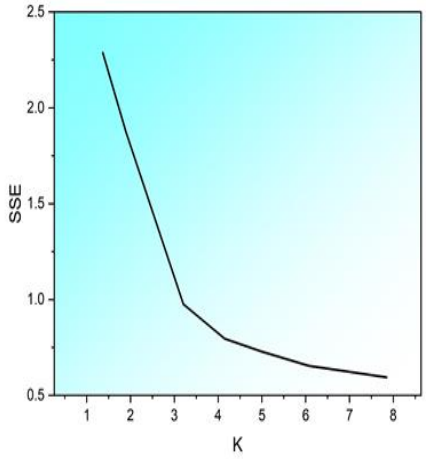

Figure 13: Number of clustering categories of the SSE index in the control group.

The forecasting of qualitative outputs and the analysis of connections among parameters are made possible by logistic regression. The purpose of this investigation is to evaluate how well the scenario clustering algorithm performs in improving the media content recommending system. A quantitative structure for assessing the algorithm's influence on suggesting pertinent information is provided by logistic regression, which advances the knowledge of the algorithm's effectiveness in the environment of new media. The results of the Logistic Regression Analysis of Media Content Recommendation are displayed in Figure 14.

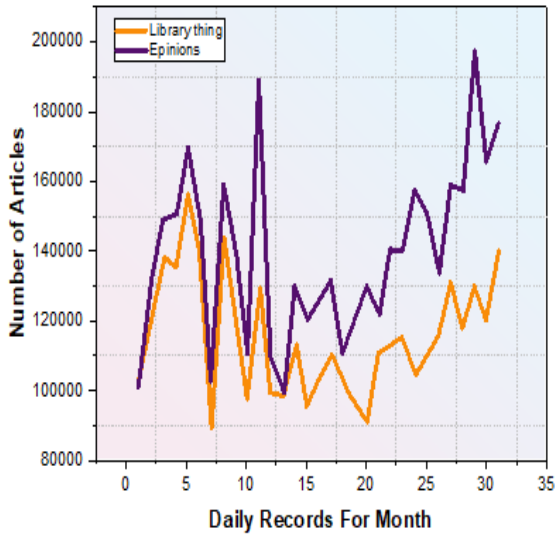

Figure 14: Logistic regression analysis results

6 Discussion

Distinct characteristics can be identified in the clustering findings of the control group, which vary between various event classifications. Category-1 events demonstrate the utmost level of participation, with a substantial number of people actively participating in discussions and a corresponding wealth of media, such as images and videos. Category-2 events are characterized by a significant level of engagement in conversations, while they have fewer supporting media aspects contrasted to Category-1 activities. Category-3 events are characterized by a significant number of individuals actively involved in conversations and a significant diversity of material, particularly images and videos. Finally, Category-4 events feature fewer individuals but make up for it with a somewhat larger use of visual media. The results indicate that there are different levels of involvement and use of media in different event classifications. This suggests that there may be subtle differences in the environment and patterns of discussions and media consumption within every classification [31]. These findings can provide insights for developing strategies to engage participants and create content for future events.

The entropy clustering technique suggested in this investigation is a new strategy to supplement conventional category classification techniques. The model of entropy clustering considers various features, such as event production, network organization, and text transportation, extensively. This is in contrast to traditional techniques that mainly focus on shallow data features and have difficulty accurately distinguishing among events due to significant information variations between categories. Employing this approach may accurately capture the subtle attributes of occurrences, enabling a more accurate measure of similarity and cluster analysis across different fields without compromising the relevancy of the information. This methodology facilitates efficient content suggestion by taking into account a wider range of characteristics and connections within the data. In conclusion, the incorporation of entropy clustering enhances the ability to identify commonalities between events, resulting in more precise categorization and improved recommendation systems that can accommodate a wide range of user preferences and requirements.

The results of the simulation experiment indicate that the scenario clustering algorithm proposed in this paper is effective and can support the analysis of media content recommendation in the new media era.

7 Conclusions

The traditional online event similarity calculation model or clustering model is subject to the constraint of surface features of events. Hence, it is challenging to build a unified similarity measurement index across the events. The continuous progress and optimization of new media technologies, have made the ways and means to access media content more abundant and diversified. Based on the scenario clustering algorithm, the business logic of the network structure and media content distribution is sorted out based on the demands of media content recommendation in this paper, and a specific clustering algorithm model is established to implement an effective representation of events. Through the comparison of the cross-event content, the media content is effectively recommended and analyzed. The results of the simulation experiment indicate that the scenario clustering algorithm proposed in this paper is effective and can support the recommendation analysis of media content in multiple dimensions, to provide high-quality media services to users. Moreover, the ever-evolving structure of new media networks and fast-evolving content environments provide continuous difficulties in developing efficient and user-friendly recommending algorithms. The future prospects for "Media Content Recommendation in the New Media Era" involve the progression of algorithmic information, the incorporation of user feedback techniques, the enhancement of privacy protections, and the utilization of emerging methods such as AI and machine learning

to personalize recommendations and enhance user experiences.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest

Funding Statement

This study did not receive any funding in any form.

References

[1] Jiang L , Yang C C . User recommendation in healthcare social media by assessing user similarity in heterogeneous network[J]. *Artificial Intelligence in Medicine*, 2017, 81(9):63-77.

[2] Yu Z , Wang C , Bu J , et al. Friend recommendation with content spread enhancement in social networks[J]. *Information Sciences*, 2015, 309(3):102-118.

[3] Liu C L, Chen Y C. Background music recommendation based on latent factors and moods[J]. *Knowledge-Based Systems*, 2018, 159(1):158-170.

[4] Daphne, Reinau, Christoph, et al. Skin Cancer Prevention, Tanning, and Vitamin D: A Content Analysis of Print Media in Germany and Switzerland. [J]. *Dermatology*, 2016,4(2):1-8.

[5] Rehman F, Khalid O, Madani S A . A comparative study of location-based recommendation systems[J]. *Knowledge Engineering Review*, 2017, 32(3):1-9.

[6] Song H , Moon N . Eye-tracking and social behavior preference-based recommendation system[J]. *Journal of Supercomputing*, 2019, 75(4):1990-2006.

[7] Sermpezis P , Spyropoulos T , Vigneri L , et al. Femto-Caching with Soft Cache Hits: Improving Performance through Recommendation and Delivery of Related Content[J]. *IEEE Journal on Selected Areas in Communications*, 2018, 4(99):1-8.

[8] Middleton S E, Krivcovs V . Geoparsing and Geosemantics for social media: Spatiotemporal Grounding of Content Propagating Rumors to Support Trust and Veracity Analysis during Breaking News[J]. *ACM Transactions on Information Systems*, 2016, 34(3):1-26.

[9] Ma J, Li G, Zhong M, Zhao X, Zhu L, Li X. LGA: latent genre aware micro-video recommendation on social media. *Multimedia Tools and Applications*, 2018,77:2991-3008.

[10] García-Perdomo V, Salaverría R, Brown DK, Harlow S. To share or not to share: The influence of news values and topics on popular social media content in the United States, Brazil, and Argentina. *Journalism studies*, 2018,19(8):1180-201.

[11] Narangajavana Kaosiri Y, Callarisa Fiol LJ, Moliner Tena MÁ, Rodríguez Artola RM, Sánchez García J. User-generated content sources in social media: A new approach to explore tourist satisfaction. *Journal of Travel Research*,2019,58(2):253-65.

[12] Ashraf M, Tahir GA, Abrar S, Abdulaali M, Mushtaq S, Mukthar H. Personalized news recommendation based on multi-agent framework using social media preferences. *In2018 International Conference on Smart Computing and Electronic Enterprise* (ICSCEE) 2018, 1-7. *IEEE*.

[13] Stubb C, Colliander J. "This is not sponsored content"–The effects of impartiality disclosure and e-commerce landing pages on consumer responses to social media influencer posts. *Computers in Human Behavior*, 2019,98:210-22.

[14] A. C , Faleye, A. A , et al. Media Portrayal of Teen Suicide: A Narrative Content Analysis of Netflix Series 13 Reasons Why[J]. *Science of the Total Environment*, 2019,3(5):19-27.

[15] Minson S , Mukerji M , Rankine J . G26Social media support for parents and young people with food allergy – an analysis of facebook content[J]. *Archives of Disease in Childhood*, 2016, 101(1): 1-19.

[16] Bach N X , Hai N D , Phuong T M . Personalized recommendation of stories for commenting in forum-based social media[J]. *Information Sciences*, 2016, 352(3):48-60.

[17] Ohsawa T . Symmetry and Conservation Laws in Semiclassical Wave Packet Dynamics[J]. *Journal of Mathematical Physics*, 2015, 56(3): 103-110.

[18] Chakrapani S K , Barnard D J , Dayal V . Influence of fiber orientation on the inherent acoustic nonlinearity in carbon fiber reinforced composites[J]. *Journal of the Acoustical Society of America*, 2015, 137(2):617-627.

[19] Dong E K , Su C P , Dong I Y , et al. Enhanced critical heat flux by capillary driven liquid flow on the well-designed surface[J]. *Applied Physics Letters*, 2015, 107(2):1004-1010.

[20] Mayara, V, Damasceno, et al. Effects of resistance training on neuromuscular characteristics and pacing during 10-km running time trial[J]. *European Journal of Applied Physiology*, 2015,3(2):1-9.

[21] Rawshan F , Park Y . Fault-tolerable and SLA-supportive architecture for TWDM-PON systems[J]. *Photonic network communications*, 2015, 30(2):143-149.

[22] Reeves S L , Fullerton H J , Dombkowski K J , et al. Physician attitude, awareness, and knowledge regarding guidelines for transcranial Doppler screening in sickle cell disease.[J]. *Clinical Pediatrics*, 2015, 54(4):336-345.

[23] Jiao Y , Liu Z , Victora R H . Renormalized anisotropic exchange for representing heat assisted magnetic recording media[J]. *Journal of Applied Physics*, 2015, 117(17):2417-2432.

[24] Toyoura K , Ohta M , Nakamura A , et al. First-principles study on phase transition and ferroelectricity in lithium niobate and tantalate[J]. *Journal of Applied Physics*, 2015, 118(6):103-110.

[25] Huang T J . Acoustic tweezers: Manipulating particles, cells, and fluids using sound waves[J]. *Journal of the Acoustical Society of America*, 2015, 137(4):2222-2229.

[26] Tinakiche N , Annou R . Oscillating two-stream instability in a magnetized electron-positron-ion plasma[J*]. Physics of Plasmas*, 2015, 22(4): 101-110.

[27] Song H , Moon N . Eye-tracking and social behavior preference-based recommendation system[J]. *The Journal of Supercomputing*, 2019, 75(4):1990-2006.

[28] Zhang Y , Zhang L , Gai S , et al. Cloning and expression analysis of the R2R3-PsMYB1 gene associated with bud dormancy during chilling treatment in the tree peony (Paeonia suffruticosa)[J]. *Plant Growth Regulation*, 2015, 75(3):667-676.

[29] Mcclure J , Morton C , Yarusevych S . Flow development and structural loading on dual step cylinders in laminar shedding regime[J]. *Physics of Fluids*, 2015, 27(6):477-539.

[30] Kieselmann J , , Rosselet A , , Scheib S , , et al. SU-E-J-118: A Systematic Analysis of Rigid Image Registration Using Patient CTs and Simulated Setup Images with a Unique Gold Standard Registration[J]. *Medical Physics*, 2015, 42(6):3291-3292.

[31] Xu Y , Zhang H , Gao H , et al. Preference discovery from wireless social media data in APIs recommendation[J]. *Wireless Networks*, 2021,5(4):1-8.