Hindawi
Mobile Information Systems
Volume 2022, Article ID 7964391, 9 pages
https://doi.org/10.1155/2022/7964391

Research Article

Big Data Analysis Application in News Service Mode Based on Genetic Algorithm

Jie Li [1] and Qiuli Wu [2]

[1] *Propaganda Department, Cangzhou Normal University, Cangzhou, Hebei 061001, China*
[2] *International Exchange Center, Cangzhou Normal University, Cangzhou, Hebei 061001, China*

Correspondence should be addressed to Jie Li; lijie1989@caztc.edu.cn

Received 2 August 2022; Revised 13 September 2022; Accepted 23 September 2022; Published 6 October 2022

Academic Editor: Shadi Aljawarneh

Due to the development and popularization of technologies such as the Internet, traditional forms of media have been greatly affected. Therefore, the way people receive news has also undergone certain changes. Due to the dramatic increase in the amount of information data and the variety of sources that are very rich, the complexity of the rational use of data information is also increasing. In order to improve the speed and accuracy of text classification, for the problems of classification efficiency and classification accuracy, this study adopts a large-scale text classification method based on genetic algorithm optimization. After designing the experimental analysis, the optimized genetic algorithm can be used to classify effectively, thus providing a new processing idea and method for the news application in the context of big data. First, we analyze and review excellent data work at home and abroad, emphatically analyze the impact of new news models on traditional news production, and propose that in the current news service model, big data can be combined to realize informatization applications. To predict information in the context of current data results, the application and sharing of data information can be realized by designing an open-source news database. The finished design of the media database can be used as the basis for the development of self-news and can take the lead in the news competition. This work analyzes and studies big data technology and genetic algorithms and introduces them into the field of news service model design, thus promoting the development of new news models.

1. Introduction

At the quantitative level, information data are constantly expanding with the rapid development of computer technology. As such information data enriches people's lives, its own complexity is gradually increasing, making it more difficult for information data to be used by people, causing people to spend a lot of money, time, and energy, resulting in a huge waste of human and material resources [1, 2]. In addition, there are a lot of useless and harmful information in the information data, which have great negative influence and adverse events on the information processing process. Therefore, how to use information data efficiently and effectively has become a hot research topic in all walks of life and fields [3]. The resources created by big data technology have made an impact on the traditional news media industry, making it undergo revolutionary changes and transformations and leaving a deep imprint, changing the information reporting form of the traditional media industry to a certain extent [4]. One of the core resources of mass media is the data foundation, and the application of big data solves the problem of information scarcity in the traditional media industry. Today, the basic problem of mass media providing information services is how to extract complex original data to obtain target-deep information [5]. Mining and presenting information to the public through searching and filtering has become a major challenge in the field of news applications. In the era of big data, data journalism is a new way of reporting news [6]. In the context of big data, mobile news clients have gradually become the preferred way for mobile netizens to obtain information due to their rich resources and convenient information acquisition methods, which can provide high-quality information content. One of the key factors for the success of today's mobile news service clients is big data technology [7]. Current news clients have their own characteristics, but there are also many similarities. Today, the commonly used news clients can be roughly divided into three categories: one is the news client created by portal websites such as NetEase, Sohu, and Tencent; the other is the news client managed by traditional media such as People's Daily, Southern Weekly, and other news clients The other type is the personalized recommendation news client produced by emerging Internet companies such as Toutiao and Yidian News. This study takes the era of big data as the research background, conducts user surveys and in-depth interviews based on the optimization process of genetic algorithms, and predicts the future development trend of data news by analyzing user experience, hoping to provide a data foundation for the development and transformation of news service models [8].

2. Related Work

Research on the current news service model: the literature believes that compared with traditional media and traditional media methods, mobile media has three unique characteristics: first, the terminal is the medium, and second, the medium is the terminal; it relies more on mobile platform operations; third, mobile media in a broad sense is not only omnimedia but also a multifunctional personal center, and its finished products have different forms and are constantly changing [9]. It can be seen from the literature that the amount of data in the world is currently in a state of exponential growth. Through big data analysis, the growing competition for new productivity can be realized and provide a data foundation for it. [10] In the era of big data, there are many challenges that can impact traditional industries, and such challenges are also opportunities, which can promote the development of the industry if grasped properly. Therefore, for the news media, how to reasonably apply related technologies is a topic worthy of further study [11, 12]. Literature shows that new-age news sites have rich news resources from journalists, editors, designers, and developers who use news data. By browsing the website, users can get the information behind the data news from all over the world, learn new techniques and new methods of effective data analysis, and find out about the latest achievements in the field by searching for training and job opportunities in user-related fields [13]. The literature assumes that the integration of news customer information is presented through various media forms such as text, video, Atlas, and sound, and the news content is presented in a multiangle and multilevel stereoscopic manner, which enriches the user's audio-visual enjoyment [14]. This article takes Sohu News Client, The Paper, and Today's Toutiao as examples, analyzes the main role and influence of big data in news clients, and discusses the growth and evolution of mobile news clients in the era of big data [15]. In the research of big data based on genetic algorithms, the literature believes that the creation and consumption of technologies, products, economies, and cultures that rely on the production, dissemination, use, and absorption of relevant information make the world an organically connected whole [16, 17].

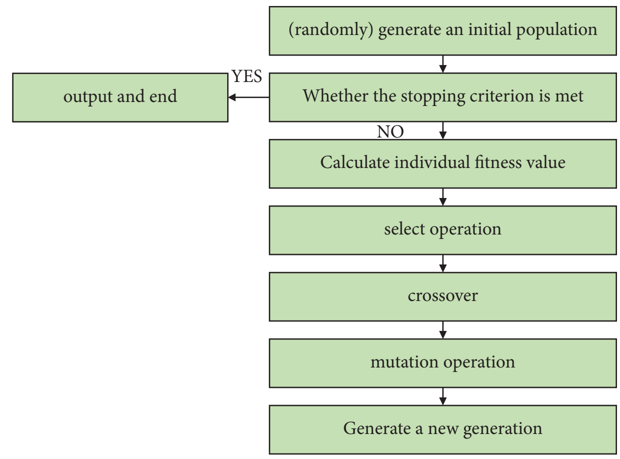

Figure 1: The calculation process of the genetic algorithm.

3. Principles of Genetic Algorithms

3.1. Theoretical Basis of Genetic Algorithm. The calculation process of the genetic algorithm is shown in Figure 1.

(1) Encode different decodings of a practical problem into bit strings, i.e., individuals (xi)

(2) Define the adaptation function f

(3) Determine the genetic strategy, crossover probability, and mutation probability

(4) Initial population

(5) Calculate the fitness $f(xi)$ of xi and then select the individual according to the individual fitness according to a certain proportion $Ps[xi]$ (the general selection strategy is the roulette algorithm)

(6) If the maximum genetic algebra is reached or the fitness requirements are met, continue to execute; otherwise, go to (5);

(7) Display the result and output

Finally, the decoding operation is performed, and the obtained solution is the optimal solution.

3.2. Design of Genetic Operators. Crossover probability is another important aspect of crossover operator design. A high crossover probability may lead to sufficient crossover in each generation, but it may also destroy a good population pattern; a lower crossover probability makes it possible to find the global optimal solution, but the speed of genetic search slows down or even stops.

In this project,

$$P_c = \begin{cases} \dfrac{P_{cmax} - P_{cmin}}{1 + \exp\left(2\left(f' - \overline{f}\right)/f_{\max} - \overline{f}\right)} + P_{cmin}, f' \geq \overline{f}, \\ P_{cmax}, f' < \overline{f}. \end{cases} \quad (1)$$

As evolution progresses, individuals tend to be more adaptable and more closely linked, resulting in individual structures becoming more and more monolithic. If evolution does not occur over many consecutive generations, it may not be possible to find an optimal solution based on the current population. At this time, the mutation rate can be increased to expand the search range. The formula is as follows:

$$P_m = \begin{cases} P_{\min} + \dfrac{t}{T_{\max}}, 0 \le \dfrac{t}{T_{\max}} \le (P_{\max} - P_{mmin}), \\ \dfrac{P_{mmax}}{P_{mmax} - P_{\min} - 1} \times \dfrac{t}{T_{\max}} + \dfrac{P_{\max}}{P_{mmin} + 1 - P_{mmax}}, (P_{mmax} - P_{\min}) \le \dfrac{t}{T_{\max}} \le 1. \end{cases} \quad (2)$$

The individuals in the genetic algorithm have no memory function, each individual can only reflect the current situation, and the scope and direction of evolution are blind, which will lead to the destruction of the stability of the population; at the same time, all individuals share information with each other in competition, so that the entire population is divided. To this end, this project introduces particle swarm optimization. Individuals in particle swarm optimization have memory functions and can control the order and direction of evolution based on historical information and current conditions. In most cases, they converge to the optimal solution faster; swarm algorithms find the optimal solution through cooperation among individuals. Only the best position in the group sends information to other particles in one direction, and individuals do not share information directly, so the entire search and update process always follows the current optimal solution, which helps maintain the diversity of the population.

3.3. Crossover and Mutation of Genetic Operators. Crossover is also called genetic recombination, and its basic principle is to recombine parts of the genes of two parental individuals to generate new individuals. Genetic recombination or crossover operation plays an important role in the genetic algorithm, which can obtain excellent new individuals. Some of the most commonly used crossover operators are selected.

3.3.1. Point Cross. The basic process of point intersection is to randomly generate one or more intersection positions and then exchange the corresponding strings of the parent individual. The basic point intersection process is shown in the following formula:

$$\begin{matrix}[01110 \mid 110] \\ [10101 \mid 001]\end{matrix} \Rightarrow \begin{matrix}[01110 \mid 001 \\ [10101 \mid 110]\end{matrix} {}^{\text{or}} \begin{matrix}[01|110|]110] \\ [10|101|001]\end{matrix} \Rightarrow \begin{matrix}[01|101|110] \\ [10|110|001]\end{matrix}. \quad (3)$$

3.3.2. Evenly Cross. The basic process of uniform intersection is to change each segment of two separate parent strings according to the probability. The specific process is to first randomly generate a control template with the same length as the parent's respective strings, 1 means exchange and 0 means no exchange, and then cross-process the parent and child strings according to the control template. For example, the generated control pattern is [10011100], and the basic process of uniform crossing is shown in the following equation:

$$\begin{matrix}[01110110] \\ [10101001]\end{matrix} \longrightarrow [10011100] \Rightarrow \begin{matrix}[11101010] \\ [00110101]\end{matrix}. \quad (4)$$

From the perspective of crossover patterns, uniform crossover and point crossover have the following characteristics: the point crossover pattern does not change much, but the search pattern is small, and the uniform crossover pattern has a high probability of changing, but the searched pattern does not change much. Therefore, in practical problems, uniform intersection is usually used when the group size is small, and point intersection is usually used when the group size is large.

The following two types of differences are usually identified:

① Basic variation

The basic process of basic mutation is as follows: randomly generate a mutation position or multiple mutation positions, and reverse the corresponding code value. The basic process of basic mutation is shown in the following formula.

$$[10101010] \Rightarrow [10001110]. \quad (5)$$

② Reverse mutation:

$$[10101010] \Rightarrow [10010110]. \quad (6)$$

4. News Big Data Processing Method Based on Genetic Algorithm

4.1. Data Sources. According to the statistical analysis of various data sources, combined with the sources of media, this study divides news data sources into the following categories:

Table 1: News distribution.

Classification category	Training text	Test text
Society	387	40
Healthy	417	47
Internationality	302	35
Finance	0	56
Film and television	0	59
Physical education	0	30
Total	1106	267

(1) Government: mainly official data released by the government and other organizations, such as government departments such as the National Bureau of Statistics of China and large organizations such as the European Union and OCED

(2) Enterprises or nongovernmental organizations: the survey data released by enterprises such as Alibaba and British Airways, as well as nongovernmental organizations such as the International Committee of the Red Cross (ICRC) and the International Accreditation Association (IPA)

(3) Universities, research institutions or individuals: universities and professional research institutions, as well as survey data or research reports published by scholars, researchers, and enthusiasts' data news

(4) Social networks: data obtained from social networks at home and abroad such as Twitter, Facebook, Weibo, and WeChat

In terms of news, excluding some repeated news, this study collects 1373 news and divides them into 6 categories, of which 1106 texts are used as training texts, and the remaining 267 texts are used as sample tests, as given in Table 1.

4.2. Acquisition of News Text Data. A sentence contains several pairs of entities and several semantic descriptions of the entities, and these semantic descriptions contain entity relationships. The correct extraction of entity pairs and their semantic descriptions from sentences is the main research goal of this section. In order to facilitate the subsequent analysis, we first define the entity pair.

An entity pair refers to two entities in a sentence that have direct dependencies. We formally define it as pair (e_i, e_j) = {<e_i, type (e_i) >, <e_j, type (e_j) >}, where e_j refers to an entity and type (e_j) represents the entity type, such as location, organization, and personnel.

An entity pair describes a sequence of segments and refers to a set of words that can describe the semantic relationship between two entities, which can be entity context. In this study, a set of gerunds in the shortest dependency path of two entity-dependent syntax trees are used as the order to describe the entities. We denote the feature sequence description of the combined entity pair as *fs* (e_i, e_j), and its definition is formally expressed as

$$fs\left(e_i, e_j\right) = \{w_i \mid pos\ \ (w_i) \in \{v, n\}, 1 \le i \le K\}. \tag{7}$$

Among them, *pos* (w_i) represents the part of speech in the vocabulary of W_i; v, n represent that the part of speech is a verb or a noun; K is the length of the segment sequence.

Heuristic rules extract sequence descriptions. According to Chinese expression habits, the closer the edit distance between words in a sentence, the stronger the semantic correlation between two words. Therefore, when taking the sequence part of an entity pair, the following rules exist:

(1) Since the dependence between words weakens with the increase of the length of the dependence distance, the words in the feature sequence of the entity pair must be words directly connected to the entity words with dependent edges

(2) If the sentence has three entities, e_1, e_2, e_3, it is determined that there are two entity pairs, pair (e_1, e_3), pair (e_2, e_3), and the dependency path dependency Path (e_1, e_3) contains dependency Path (e_2, e_3); then, the shortest dependency path between these two entities is as follows:

$$\text{shortPath}\left(\text{pair}\left(e_2, e_3\right)\right) = \text{dependencePath}\left(e_2, e_3\right), \tag{8}$$

$$\begin{aligned}\text{shortPath}\left(\text{pair}\left(e_1, e_3\right)\right) = \text{dependencePath}\left(e_1, e_3\right) \\ - \text{dependencePath}\left(e_2, e_3\right).\end{aligned} \tag{9}$$

According to the semantic localization rules, we extract entity pairs with semantic dependencies from sentences and then extract the descriptive feature sequences of entity pairs that determine the relationship through partial descriptive sequence extraction rules. It should be emphasized that if an entity pair has multiple dependency paths, each entity pair and triplet consisting of segment description sequences are considered to be different individuals, that is, multiple relationships are allowed for the same entity pair.

Since this study extracts entity pairs based on syntactic dependencies and uses the shortest dependency path as the feature sequence of entity pairs, therefore, the number of feature words contained in the feature sequence of the entity pair cannot be determined, and the similarity cannot be obtained by using a conventional calculation method. Based on this background, this study decides to use the sequence kernel function to measure the entity similarity. The calculation formula is

$$K(X, Y) = \frac{1}{Z(X, Y)} \sum_{n=1}^{K} K_n(X, Y). \tag{10}$$

Among them, X and Y represent the feature description intervals of two pairs of entities, and the lengths of the two intervals are not necessarily the same; Z (X, Y) is the normalization factor, which is defined as follows:

$$Z(X, Y) = \sqrt{\sum_{n=1}^{|X|} K_n(X, X) \times \sum_{n=1}^{|Y|} K_n(Y, Y)}. \tag{11}$$

The calculation formula of the semantic kernel function is as follows:

$$K_n(X,Y) = \sum_{u\in \sum_n} \sum_{i:u=X[i]} \sum_{j:u=Y[j]} \lambda^{l(i)+l(j)} \times \prod_{k=1}^{n} SIM\left(X_{i_k}.\text{word}, Y_k.\text{word}\right). \tag{12}$$

Among them, u represents the common sequence of two entities in the feature sequence; λ is the decay factor starting from (0, 1), using the entire text of the encyclopedia to train the word vector, and calculating the difference between the two words by the similarity of the word vector, the semantic similarity between the two is calculated as follows:

$$SIM\ (W_A, W_B) = \frac{\overrightarrow{W}_A \cdot \overrightarrow{W}_B}{\|\overrightarrow{W}_A\| \cdot \|\overrightarrow{W}_B\|}. \tag{13}$$

In summary, we use (13) to calculate the similarity of two sets of features. The disadvantage of this method is the high computational complexity, but this complexity is usually affected by the length of the feature sequence. In this study, the shortest reliable path is used as the segment description sequence, and the length itself can be well controlled, so that the operation cost will not be too high.

First, we create a similarity matrix of entity pairs by calculating the similarity between entity pairs, which is represented as follows:

$$\begin{pmatrix} k(\text{entityPair}_1, \text{entityPair}_1) & \ldots & k(\text{entityPair r}_1, \text{entityPair r}_n) \\ \vdots & \ddots & \vdots \\ k(\text{entityPair}_n, \text{entityPai r}_1) & \cdots & k(\text{entityPair r}_n, \text{entityPair}_n) \end{pmatrix}. \tag{14}$$

The general process of the spectral clustering algorithm is shown in Figure 2.

According to the description of Figure 2, combined with the requirements of this section, we summarize the detailed steps of the entity pair set spectral clustering algorithm as follows:

Step 1: Construct the similarity matrix of the sample set, measure the correlation between samples according to formula (10), construct the similarity matrix W shown in formula (12), and construct the degree matrix D at the same time

Step 2: Calculate the Laplace matrix $L = D - W$ and standardize the D matrix

Step 3: Perform feature synthesis on the matrix and select the eigenvector f corresponding to the minimum k eigenvalues

Step 4: Normalize the matrix composed of the corresponding eigenvectors f in the row and finally obtain the $n \times k1$ eigenmatrix F

Step 5 Each row of the $k1$-dimensional vector of F is a sample, a total of n samples input the clustering algorithm, cluster the samples, and finally output the clustering result.

Whether to divide the cluster centers in the X-means algorithm is determined by the division criterion. We pick a cluster center, then create a center next to it, and evaluate whether both centers perform better, if so, split the class, and if not, go back to the previous step. The algorithm judges performance by the BIC score. The BIC calculation formula is as follows:

$$BIC\left(M_j\right) = \hat{l}_j(D) - \frac{p_j}{2} \cdot \log\ R, \tag{15}$$

where D represents the dataset, M_j represents the model, and different models represent different K values.

Relation label extraction refers to extracting the corresponding words from the same category entity as the category relation label for the feature set description sequence. This method aims to obtain the most discriminative word from the feature sequence category of the entity pair as the label of the category, which is based only on the word frequency, ignoring the semantic features of each word. However, the idea of this method is simple and easy to implement, and the accuracy rate is high. In this study, DCM is used for label extraction. The plan is divided into the following two phases:

First, we determine the importance of the feature word fi in the class; we define it as $WC_{i,k}$, and the calculation formula is as follows:

$$WC_{i,k} = \frac{\log_2\left(df_{i,k} + 1\right)}{\log_2\left(N_k + 1\right)}, \tag{16}$$

where $df_{i,k}$ represents the number of entity pairs with a certain word fi in category k; N_k represents the total number of entity pairs in the dataset. The formula for calculating the importance between categories is as follows:

$$CC_i = \log \frac{N \cdot \max_{k\in C_i}\left\{WC_{i,k}\right\}}{\sum_{k=1}^{N} WC_{i,k}} \cdot \frac{1}{\log N}. \tag{17}$$

Among them, C_i is the set of all the classes with the feature word f_i, and we take the feature word with the highest weight in each class as the label of the class relationship.

$$\text{Weight}\left(f_i\right) = \frac{W_{i,k}^2 \times CC_i^2}{\sqrt{W_{i,k}^2 \times CC_i^2}}. \tag{18}$$

The words of the relationship label between the entity pairs are usually arranged in the order of the entity pair features, so after obtaining the category label, we use the formula (19) to select the feature description sequence in the word label with the highest matching degree of the category vocabulary, and put it as a specific label for entity pairs.

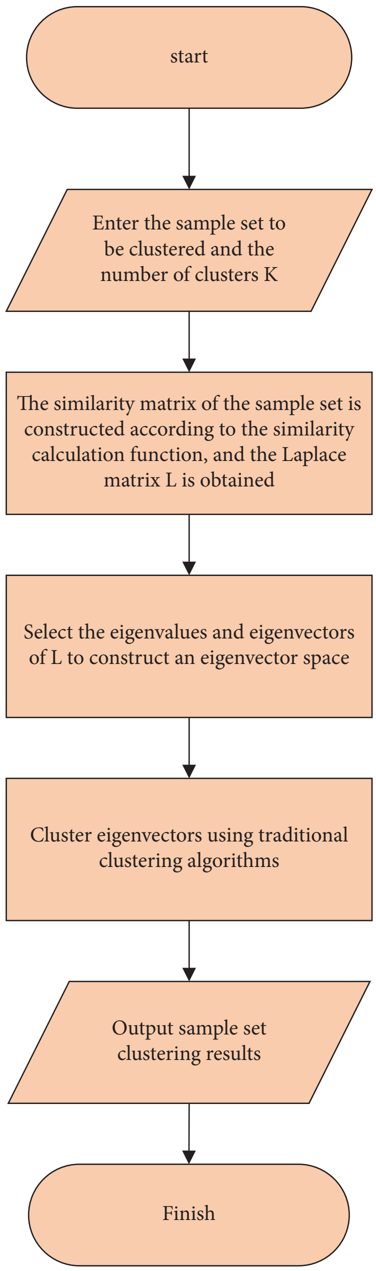

FIGURE 2: Flowchart of the spectral clustering algorithm for similar relationship entities.

$$\underset{f_i}{\arg\max}\,\mathrm{sim}\left(f_i, f_i^C\right). \tag{19}$$

Among them, $f_i \in f_s\ (e_i, e_j)$, that is, the label word is generated from the feature sequence of the entity pair pair (e_i, e_j).

4.3. Evaluation Indicators. Evaluation metrics usually include text classification complexity and text classification effectiveness. The classification types of text classification are generally divided into 4 categories, namely, c_i classification categories. One is the amount of text information that belongs to the c_i classification category and is correctly classified into the c_i classification category, which is recorded as TP; the other category belongs to the c_i classification category, but is wrongly classified. Classification. The amount of text information assigned to other classification categories is denoted as FP; one category is the amount of information in the text that does not belong to the c_i classification category, but is wrongly placed in the c_i classification category, denoted as FN; other categories do not belong to the classification category. The amount of textual information is not classified in a categorical category, but is represented as TN. The performance evaluation indicators are recall rate, precision rate, and F measurement, and the specific content is defined as follows.

4.3.1. Accuracy (P). The specific calculation formula of the accuracy rate is as follows:

$$P = \frac{TP}{TP + FP} \times 100\%. \tag{20}$$

4.3.2. Recall Rate (R). The specific calculation formula of recall rate is as follows:

$$R = \frac{TP}{TP + FN} \times 100\%. \tag{21}$$

4.3.3. F Measurement. As can be seen from the above, precision and recall are two opposite standard evaluation metrics. F measurement is a measurement method that combines the characteristics of the two. The specific calculation formula is as follows:

$$F_\beta(P, R) = \frac{\left(\beta^2 + 1\right)\mathrm{PR}}{\beta^2 P + R}, \tag{22}$$

where β is a constant. By varying different values of the beta constant, we can control and modify the effects of precision and recall. If β is 1, then F_β is equal to the value of F_1, which is the average of recall and precision. The specific calculation formula is as follows:

$$F_1 = \frac{2\mathrm{PR}}{P + R}. \tag{23}$$

4.3.4. Macroaverage and Microaverage. Through the analysis of the above three methods, these methods analyze the classification results of a classification category. When we need to evaluate a classification method, we usually average the text classification results, such as macroaverage and microaverage. However, if there are many differences in the classification categories of text information, then the values of the two are also very different.

Table 2: Comparison of text classification effects based on genetic algorithms.

	Optimization algorithms			Traditional algorithms		
Number of clusters	Recall	Precision	*F*-measure	Recall	Precision	*F*-measure
2	0.593	0.483	0.532	0.558	0.365	0.440
3	0.849	0.861	0.855	0.696	0.688	0.692
4	0.587	0.812	0.681	0.390	0.500	0.438
5	0.466	0.622	0.532	0.455	0.737	0.563
6	0.376	0.673	0.482	0.370	0.712	0.487
7	0.356	0.860	0.503	0.311	0.674	0.426
8	0.198	0.740	0.313	0.306	0.194	0.237
Mean	0.487	0.719	0.554	0.438	0.550	0.467

4.4. Analysis of News Big Data Processing Results. It can be seen from Table 2 that the precision, recall, and F-measure are greatly improved after text classification is combined with the genetic algorithm. It can also be seen in the analysis of the number of clusters that the number of clusters is defined as 3 clusters. The group effect has the largest measure *F* in groups 2–8. Therefore, we can clearly see that after grouping the text information that cannot be classified into existing categories, the grouped categories of the text information can be effectively identified. Grouping categories can be added to existing categories through an optimization algorithm. Textual information that may not be classified into existing categories is processed in the future.

This study also compares some commonly used methods for clustering similar textual information, aiming to divide the groups into 3 groups. As can be seen from Figure 3, in the case of similar text information, the recognition effect of the method in this study is much greater than that of other algorithms.

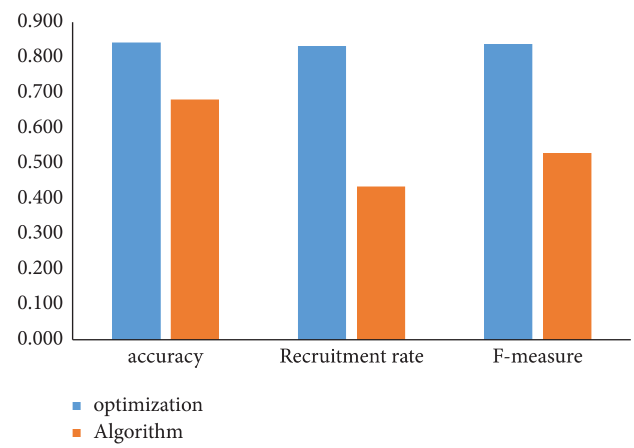

Figure 3: Comparison of the effect of the algorithm in this study and some existing algorithms.

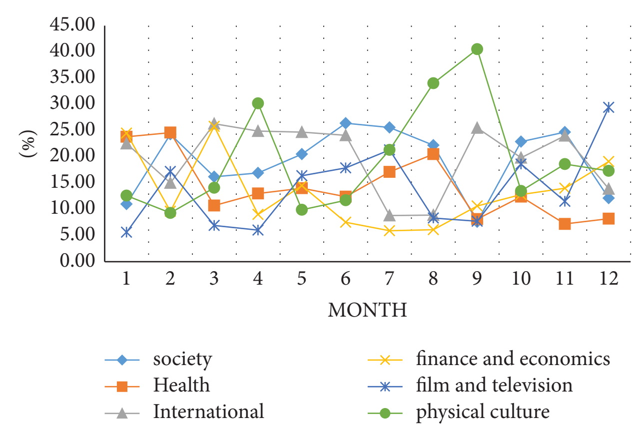

Figure 4: Changes in different types of news.

5. The Changing Trend of the Current News Service Model

5.1. Changes in News Service Content. Compared with film and television, sports, and financial news, social news and international news are more inclined to hard news. Choosing this topic reflects setting the mobile news client's agenda to focus on news content coverage. At the same time, public use of mobile news media is often limited to a fraction of the time. Users expect short, timely, and rich news content and high-quality news coverage in a very short period of time. The data show that during the morning peak hours of 08: 00 to 10: 00 am, viewers tend to get high-quality hard news.

As can be seen from Figure 4, the selection of different types of news topics is uneven in the sampling time series of the mobile news client: on the one hand, in the reporting of social news, political news, and military news, the three show similar situations. On the other hand, the other three types of news are quite different, showing an irregular state. The reason is that some news topics have their own "soft" attributes, that is, movies, sports, and finance, and other news types with soft news characteristics have relatively low timeliness requirements, and client-side publishers will "prioritize" to browse high-frequency news. Hard news with a large amount of information and a high degree of topicality are chosen to deliver. We should take advantage of the number of reports, complete the agenda setting, and guide the audience to pay attention to the news content with a large proportion of reports. On the other hand, in the morning peak hours of mobile news browsing, the audience's tendency to choose news types also affects and determines the market and value of hard news, thus diluting the existence of soft news.

5.2. Analysis of News Service Demand and Satisfaction. As shown in Figure 5, in the mobile news accuracy market survey, it can be seen that only 1.89% of people are dissatisfied with mobile news; in addition, 53.56% of users are satisfied with the accuracy of mobile news, the highest proportion. This shows that, as the core competitive push service of mobile news, it can generally better meet the needs of users.

As shown in Figure 6, according to the market survey questionnaire, when 50% of users were asked "what makes you most satisfied with the news model in the new era," they said that mobile news is most satisfied with "abundant information," some in-depth interviewees indicated that mobile news is rich and comprehensive, and the ability to refresh new news at any time is its most satisfying advantage. This indicates that users have specific needs for the quantity and richness of news content.

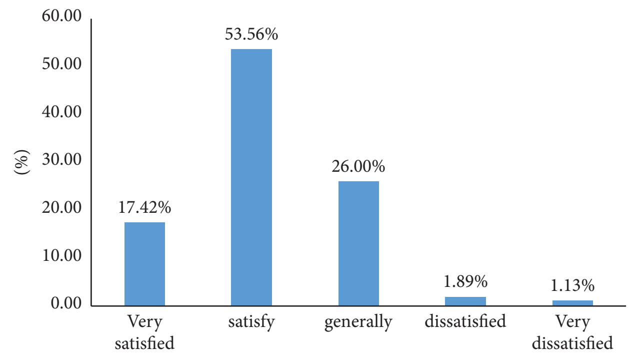

Figure 5: User satisfaction of news mode in the new era.

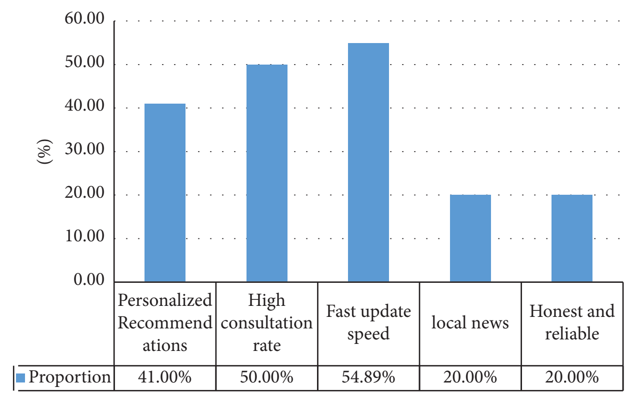

Figure 6: User satisfaction points of the new era news model.

5.3. Development Trend of Current News Service Mode

5.3.1. Predict Information Based on Existing Data Results. Predictive journalism is forward-looking reporting of events that are about to happen but have not yet happened. It focuses on the scientific prediction of the development process or news viewpoints. Traditional predictive news reports emphasize the subjective perception of reporters, but due to factors such as insufficient personal experience and limited knowledge, predictions often lack scientificity and accuracy. The lack of scientific and quantifiable forecasting methods and technical analysis methods hinders the accuracy and objectivity of forecasting reports to a certain extent.

5.3.2. Design of Open-Source News Database, Application, and Sharing of Data Information. As data technology continues to mature, more databases appear. These databases are diverse and include not only publicly available government information but also databases created by third parties, as well as the work of media organizations. These databases have various forms, some are for profit and require users to pay for use and some are open source and free. The raw data used in data journalism are open and free.

5.3.3. Design of the Media Database as a Competitive Advantage for Its Own News Development. Since the rapid development of Internet technology, many media have created their own information databases to adapt to the changing environment. According to new communication technologies, in the era of big data, different environments have different applications for big data. Enterprises form the foundation of big data by establishing their own databases. However, newspaper big data are not the same as news data. It is only a small part of the application, including the source product.

6. Conclusion

With the continuous development of the information age, the amount of data is increasing day by day, which has brought a profound imprint on the news industry, and even changed the news reporting methods and ecology of traditional media. With the development of the times, the types of news data will continue to improve, and the demand will continue to grow. The news industry needs to push news according to the needs of the audience and improve the news service model for optimization and development. According to the above data, media products can be targeted, accurately analyze the characteristics of each user, and push on-demand content in a targeted manner, thereby winning clicks and user satisfaction. Although the use of new technology can make up for the shortcomings of traditional news production and is an innovative technology, it does not mean blindly following technological progress. News services must still follow industry rules and maintain the original intention of the media.

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 that they have no conflicts of interest.

References

[1] C. Elmer, F. Badenschier, and H. Wormer, "Science for everybody? How the coverage of research issues in German newspapers has increased dramatically," *Journalism & Mass Communication Quarterly*, vol. 85, no. 4, pp. 878–893, 2008.

[2] X. Liu and Z. Li, "False data attacks against AC state estimation with incomplete network information," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2239–2248, 2017.

[3] X. Liao, Y. Li, and B. Lu, "A model for selecting an ERP system based on linguistic information processing," *Information Systems*, vol. 32, no. 7, pp. 1005–1017, 2007.

[4] M. T. Odenkirk, D. M. Reif, and E. S. Baker, "Multiomic big data analysis challenges: increasing confidence in the interpretation of artificial intelligence assessments," *Analytical Chemistry*, vol. 93, no. 22, pp. 7763–7773, 2021.

[5] S. Koike, S. Yamaguchi, Y. Ojio, K. Ohta, and S. Ando, "Effect of name change of schizophrenia on mass media between 1985 and 2013 in Japan: a text data mining analysis," *Schizophrenia Bulletin*, vol. 42, no. 3, pp. 552–559, 2016.

[6] E. Borges-Rey, "Unravelling data journalism: a study of data journalism practice in British newsrooms," *Journalism Practice*, vol. 10, no. 7, pp. 833–843, 2016.

[7] C. Pentzold, C. Brantner, and L. Fölsche, "Imagining big data: illustrations of "big data" in US news articles," *New Media & Society*, vol. 21, no. 1, pp. 139–167, 2019.

[8] R. G. Picard, "Twilight or new dawn of journalism? Evidence from the changing news ecosystem," *Journalism Practice*, vol. 8, no. 5, pp. 488–498, 2014.

[9] Y. Aoudni, C. Donald, A. Farouk et al., "Cloud security based attack detection using transductive learning integrated with Hidden Markov Model," *Pattern Recognition Letters*, vol. 157, pp. 16–26, 2022.

[10] T. He, C. Huang, M. Li, Y. Zhou, and S. Li, "Social participation of the elderly in China: the roles of conventional media, digital access and social media engagement," *Telematics and Informatics*, vol. 48, Article ID 101347, 2020.

[11] A. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Boldú, "A review on the practice of big data analysis in agriculture," *Computers and Electronics in Agriculture*, vol. 143, pp. 23–37, 2017.

[12] F. Arena and G. Pau, "An overview of big data analysis," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1646–1653, 2020.

[13] C. K. S. Leung, "Big data analysis and mining," in *Advanced Methodologies and Technologies in Network Architecture, mobile Computing, and Data Analytics*, D. B. A. Mehdi Khosrow, Ed., pp. 15–27, IGI Global, 2019.

[14] J. Zhang, X. Yang, and D. Appelbaum, "Toward effective big data analysis in continuous auditing," *Accounting Horizons*, vol. 29, no. 2, pp. 469–476, 2015.

[15] A. Kim and A. R. Dennis, "Says who? The effects of presentation format and source rating on fake news in social media," *MIS Quarterly*, vol. 43, no. 3, pp. 1025–1039, 2019.

[16] D. Vazquez, C. Dennis, and Y. Zhang, "Understanding the effect of smart retail brand—consumer communications via mobile instant messaging (MIM)—an empirical study in the Chinese context," *Computers in Human Behavior*, vol. 77, pp. 425–436, 2017.

[17] T. A. Mohammed, O. Bayat, O. N. Uçan, and S. Alhayali, "Hybrid efficient genetic algorithm for big data feature selection problems," *Foundations of Science*, vol. 25, no. 4, pp. 1009–1025, 2020.