

## *Retraction*

# **Retracted: The Identification and Dissemination of Creative Elements of New Media Original Film and Television Works Based on Review Text Mining and Machine Learning**

### **Mathematical Problems in Engineering**

Received 3 October 2023; Accepted 3 October 2023; Published 4 October 2023

Copyright © 2023 Mathematical Problems in Engineering. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] X. Yu, "The Identification and Dissemination of Creative Elements of New Media Original Film and Television Works Based on Review Text Mining and Machine Learning," *Mathematical Problems in Engineering*, vol. 2022, Article ID 5856069, 9 pages, 2022.

## Research Article

# The Identification and Dissemination of Creative Elements of New Media Original Film and Television Works Based on Review Text Mining and Machine Learning

Xiaonan Yu 

Deputy Director of Graduate Office, Shandong College of Arts, Jinan, Shandong 250300, China

Correspondence should be addressed to Xiaonan Yu; z00085@sdca.edu.cn

Received 23 April 2022; Revised 16 June 2022; Accepted 18 June 2022; Published 16 July 2022

Academic Editor: Zaoli Yang

Copyright © 2022 Xiaonan Yu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous growth of the film market and the consumption demand of audiences, the value of film content has become increasingly apparent. Extracting movie content elements is an important step in quantitative analysis of movie content. In this paper, based on text mining technology, deep data analysis of movie reviews is carried out using TF-IDF and machine learning to visualize high-frequency words, and sentiment analysis is performed on reviews to find out the hidden deep information behind the big data of movie reviews. Although the extracted keywords can draw the content and characteristics of the film for us, they cannot establish the correlation with the creative elements of the film and television works. Therefore, in this paper, the extracted keywords are clustered to find the central words of these words. First of all, in order to improve the representativeness of keywords, the Epoch data set and the Batchsize data set are used to conduct experimental analysis on the model in this paper. Through comparative experiments, it is concluded that when Batchsize = 32 and Epoch = 25, the model achieves the optimal classification effect. The analysis shows that when the training times are too small, the model is not fully learned, and when the training times are too many, the model will be overfitted, which will reduce the accuracy of the model, resulting in the opposite effect. The keywords extracted above can draw the content and characteristics of the movie for us.

## 1. Introduction

Based on the current new media environment, the creation of new media original film and television works cannot only be limited to the orientation of control and prediction, nor can it only unilaterally pursue the meaning understanding and interpretation of the works, but should focus on real-time and comprehensive changes to the needs of the audience, analysis, and more precise interpretation [1]. Provide decision-making reference for producers of original film and television works in content planning, and provide strong support for new media film and television from “creation-led” to “demand-led,” leading to cultivate the audience’s aesthetics, while penetratingly changing the public’s way of thinking. In recent years, due to the increase in production costs of TV dramas and movies, copyright fees have been soaring, and major video websites and portals have invested

heavily to develop the market for original film and television works with unprecedented attitude and fighting spirit [2]. Because of the high cost of purchasing copyrights, video sites are looking for ways to reduce their costs. Although the broadcasting platform of original film and television works is primarily a video website, there is no strict distinction between the production and broadcast of original film and television works. At present, the production and broadcasting of original film and television works presents a variety of cooperation modes [3]. Some original film and television works are invested by the video website and produced by professional institutions, and the video website is the sole copyright owner. Some original film and television works are separated from production and broadcast; that is, the video website only provides a broadcast platform and does not participate in the investment and production of the play, but shares the broadcast revenue with the investors and

producers of the online film and television drama through the resources of the broadcast platform. Some original film and television works implement the integration of production and broadcasting. Video websites not only invest by themselves, but also set up production teams to produce by themselves.

Text sentiment analysis has always been a very important topic in the field of natural language processing. In today's era of the full popularity of the Internet, information on major social media platforms has exploded, and movie reviews contain important information about the true evaluation of movies. This provides a public opinion on the progress of the film industry. Movie viewers on the Internet usually write down their true feelings about the movie, and other movie viewers learn about the movie by viewing the movie reviews and choose whether to watch it [4]. This paper provides massive data for the topic of text sentiment analysis. This paper focuses on two aspects of text sentiment analysis, text sentiment classification, and Chinese word segmentation. This paper conducts in-depth research on representative technologies in the process of technological development, focusing on machine-based learned text sentiment analysis techniques, and made induction and combing. The recurrent neural network is selected as the key research model in this paper, and the corresponding model improvements are made according to the characteristics of the two scenarios of text sentiment classification and Chinese word segmentation.

The creative elements of film works are an important feedback on the level of film creation and production. Audience word-of-mouth and movie viewing experience have become more and more obvious in promoting and promoting film elements, which is one of the most significant and important changes in the film market in recent years [5]. After the rapid development of the number of movies and the construction of theaters, movies are no longer a fashionable and trendy way of consumption, and the audience's time is fragmented; the time cost is getting higher and higher, coupled with the diversified development of movie viewing channels. The consumer demand of the audience is also changing, turning to the requirements for the quality of the film, and the value of the film content itself is gradually becoming more prominent. Therefore, the audience's preferences for movies and what kind of movies are more popular are the key concerns of movie creators, investors, distributors, theater chains, and theaters. In-depth analysis of movie content has also become more and more important. The content element is an important step in quantitative analysis of film content [6]. In this paper, two methods of review text mining and machine learning are combined to obtain elements that can represent movie content, and the obtained movie content elements and movie box office are modeled and analyzed. The results show that the two are strongly correlated. In order to realize the clustering of words, we first construct a corpus space for all the comments and obtain each word and the words that frequently appear in its context through a large number of sampling. Then, each part of speech of each movie is regarded as a separate word space. Words with similar

contexts in the original corpus space have similar Euclidean distances for the vectors in the projected space. From this, we can quantify the similarity between words to achieve clustering.

The innovative contribution of this paper includes the application of text mining technology and machine learning algorithm to the recognition of film elements. Through descriptive analysis and affective analysis, we try to find the useful information hidden behind the user element recognition. In the description and analysis, the method of calculating the probability of high-frequency words and generating graph cloud is used to express high-frequency words intuitively. In the context of big data, in-depth analysis of the audience's element identification is conducive to an objective and comprehensive evaluation of the film, reflect the audience's real feelings, and help other viewers make decisions about whether to watch the film, which has a certain practical significance.

Chapter arrangement of the article: the research is divided into five parts. The first part introduces and summarizes the research of relevant scholars in power element recognition, machine learning, and text mining. The second part introduces the relevant literature and analyzes the correlation between words to determine the semantic tendency of target words. The third part makes a comparative experiment with Epoch parameters and Batchsize parameters. The fourth chapter is the experimental analysis and verification, which expounds the experimental results of Epoch parameter comparison experiment, batch parameter comparison experiment and emotional classification of film and television text. Finally, the full text is summarized.

## 2. Related Work

Today, with the rapid development of the film and television industry, the way of dissemination of film and television works fully reflects the views of freedom of speech and everyone's participation in the new era. In this regard, how to obtain and analyze consumers' emotional tendencies is very important. By designing a review sentiment analysis based on machine learning, the sentiment tendency analysis is carried out on the large-scale film and television review information text, the emotion contained in the text is analyzed, and the opinion of the film and television review text on the film and television is judged, which is helpful for helping users, film and television producers, and it has certain practical significance and application value for the issuer to grasp more comprehensive film and television screening information and make decisions. The film content contains many elements. In order to summarize the film content as much as possible and to be able to analyze the correlation between the film content elements and the box office revenue, the elements of the film content should be determined first.

Xiao and Zhao extended the range of polar words to nouns. They hand-selected seed words and used an iterative method to obtain noun evaluation phrases [7]. Lisen proposed the method of point mutual information to judge the tendency of praise and derogation of words. First, the seed

words were selected, and the semantic distance between other words and the seed words was calculated to determine the praise and derogation of words. This method is more dependent on the selection of seed words, but it can be applied to the evaluation and discrimination of various parts of speech, and the scope of application is large [8]. Ravi et al. extract a feature set from the corpus and determine the semantic tendency of the vocabulary by analyzing the relationship between the feature set and the marked text [9]. Scholderer et al. discriminated tendentious words and collocation relationships based on the level annotation of the corpus [10]. Zhu et al. took the positive and negative words in GI (General Inquirer) and WordNet as seed words, obtained an expanded large-scale emotional word set, and used this as a classification feature, using the machine. The learning method automatically classifies the positive and negative meanings of texts [11]. Short uses “good” and “bad” as the seed words of sentiment words and uses WordNet to calculate the semantic distance between the new word and the two words, and this semantic distance is used to determine the tendency of words [12]. Bruynooghe et al. proposed a latent variable model that can effectively analyze the semantic tendency of phrases [13]. In order to further improve the efficiency of text sentiment classification, Samizade and Abad extract strong collocations from documents as seed words and achieve good classification results [14]. Li et al. obtain the propensity of the target vocabulary by calculating the similarity between the target vocabulary and the labeled vocabulary in HowNet, and on this basis, they propose a vocabulary semantic propensity calculation method based on semantic similarity and semantic correlation field [15]. Yadollahi et al. use the vocabulary set in HowNet as the seed word and determine the semantic tendency of the target word by calculating the correlation between the target word and the seed word [16]. Deotale et al. proposed the polar coordinate method of word tendency and used the balanced mutual information method to calculate the word tendency [17]. Popoff et al. proposed a sentence-level sentiment classification method based on tree kernel function by combining the syntactic features and dependency features of sentences and other structural features and plane features and achieved good results [18]. Through syntactic analysis, Huang et al. identify sentence evaluation subjects and their attributes and calculate topic semantic tendency based on their attributes and description relationships [19]. Kaur and Kumar used machine learning methods for the first time in chapter-level sentiment classification. They represented text as n-grams and found that the unigram feature representation method achieved better classification results. They also compared Naive Bayes, maximal descendants, and support effects of vector machine methods on text sentiment classification performance [20]. Xiao et al. selected adjectives, adverbs, and verbs as emotional words to classify texts and achieved ideal results. Gamon first fuses features and then performs feature selection to achieve higher accuracy on noisy corpora [21]. Taniguchi and Tsuda first used spectral analysis techniques to mine unambiguous subjective reviews with a single sentiment polarity and then classified the ambiguous reviews

through a combination of active learning and ensemble learning [22].

Video culture is the product of the consumer society. It has a wide and far-reaching impact on modern society through industrial mass production and various mass media. Although humans have long had visual experience, today, vision has been improved to an unprecedented height. This paper aims to improve the performance of text sentiment analysis system, mainly studies how to apply deep learning technology in text sentiment analysis and how to introduce various improvement mechanisms on the basis of the original model to achieve the purpose of improving analysis performance. The improvement method proposed in this paper will help to improve the performance of the application system based on text sentiment analysis, so it has certain scientific research significance and application value.

### 3. Sentiment Classification Method of Film and Television Works

*3.1. Machine Learning.* The process of sentiment classification based on machine learning can be roughly divided into two parts, one part is the learning process, and the other part is the emotion classification process. Among them, the learning process includes a training process and a testing process. During the training process, the training set is trained to obtain a classifier, which is used to classify the sentiment of the test set, and the test results are fed back to the classifier to further improve the training method and generate new classifications. Finally, the final generated classifier is used to classify the sentiment of the new text [23]. The general process is shown in Figure 1.

Text preprocessing is the first step in text sentiment classification, and the quality of the preprocessing results directly affects whether future analysis and processing can be carried out smoothly. The purpose of text preprocessing is to standardly extract the main content from the text corpus and remove the information that is irrelevant to text sentiment classification. For Chinese preprocessing, its main operations include standard encoding, filtering illegal characters, word segmentation, and removing stop words. As far as the film itself is concerned, there seems to be a shift from a language center to a visual center. Visual art is based on film as a carrier, and itself is composed of two aspects: language and image. In the early films, literature accounted for a large proportion [24].

The text contains many words with parts of speech such as auxiliary words and function words, as well as high-frequency words that often appear in the text, but they have little meaning for sentiment classification. These words are collectively referred to as stop words. There are generally two ways to construct the stop word list, manual method or automatic statistics by machine. The existence of stop words will not only increase the storage space, but also may form noise and affect the accuracy of sentiment classification. Therefore, it is necessary to filter stop words in the text.

After preprocessing the text and expressing it formally, a high-dimensional and sparse feature space is obtained, and the number of features can reach tens of thousands or even

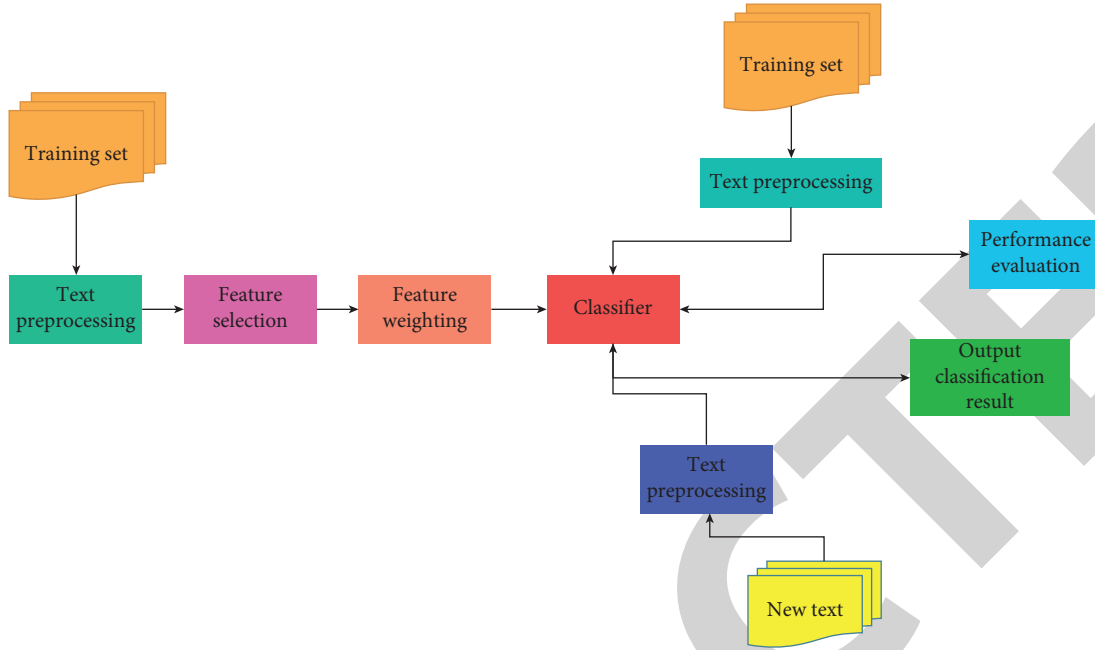


FIGURE 1: Flowchart of sentiment classification of machine learning method.

hundreds of thousands. Such a high-dimensional feature space will have a considerable impact on the classification of text, which not only makes the calculation time longer, but also reduces the classification accuracy to a large extent. Many classifiers cannot adapt to the high-dimensional feature space, and many features in the high-dimensional space are not beneficial to the classification, and many may even form noise, which greatly reduces the classification performance. According to relevant research studies, the features in the feature set have different contributions to the classification, and only a small part of the features have positive significance for the classification. Feature selection is to select a small part of the features from the original high-dimensional feature set as the classification features of the classifier. Obviously, how to select the feature subset that best represents the text category from the original feature set is what the feature selection of text needs to study.

**3.2. Comment Text Mining.** In recent years, affective analysis has become a hot topic in the field of natural language processing and has been widely used in many natural language processing problems. According to the granularity of text processing, affective analysis can be divided into word level, phrase level, and sentence level. The descriptive analysis of film review text can describe film review to a certain extent, but it cannot understand the emotional tendency behind the review. In order to understand the deep semantics of film review text, it is necessary to analyze the emotion of film review, use machine learning algorithm to determine whether the emotion expressed by film review is positive or negative, and try to find the theme behind film review. It combines text mining and natural language processing.

The review text mining method is a classification method based on statistics, which has been widely used in text

classification because it is a simple and effective linear classification method. The review text mining classifier is based on the assumption of feature independence; that is, it is assumed that the influence of an attribute on a given category is independent of other attributes. On the one hand, this assumption greatly reduces the computational complexity of the review text mining classifier; on the other hand, it also makes its classification effect less stable. Its calculation formula is

$$P(t, c) = \frac{1 + \sum_1^D N(t, d)}{v + \sum_{s=1}^D \sum_{t=1}^D N(t, d)}, \quad (1)$$

where  $D$  is the total number of training texts,  $N$  is the word frequency in the feature, and  $v$  is the total number of features. The review text mining method is an instance-based classification method, and it is also an earlier machine learning method applied to text classification. The core idea is as follows: for a given new text, examine the  $k$  texts that are closest to the text in the training text set, and determine the category of the new text according to the category to which the  $k$  texts belong. In the review text mining classification method, the value of  $k$  is very important. If the value of  $k$  is too small, it cannot fully compare each category with the text to be classified; and if the value of  $k$  is too large, the set of candidate texts will be too large, resulting in noise increases, affecting classification performance. There is no good method for the selection of  $k$  value. Generally, an initial value is adopted first, and then the value of  $k$  is gradually adjusted according to the experimental results. The cosine similarity formula is generally used to calculate the similarity, as shown in

$$\text{sim}(d, d) = \frac{\sum_{k=1}^M t \times t}{\sqrt{\sum_{k=1}^M t^2 \times \sum_{k=1}^M k^2}} \quad (2)$$

Among them,  $d$  is the text to be classified,  $M$  is the dimension of the vector, and  $t$  is the feature weight. The review text mining model provides the output layer with complete past and future context information for each point in the input sequence by scanning the input sentence from front to back and from back to front, and the hidden layers of the review text mining are all connected to the same output layer, the context information is fully utilized, as shown in Figure 2.

Any combination of any linear operations is still itself a linear operation, which means that any multilayer perceptron with multiple linear hidden layers is completely equivalent to any other multilayer perceptron with a single linear hidden layer. This is in stark contrast to nonlinear networks, which can play a considerable role by using successive hidden layers to rerepresent the input data. Data preprocessing removes incomplete and inconsistent data and excludes low-quality data. For example, each data content has 5 dimensions, and one or more of the single data are omitted for removal. Read the crawling data information, slice and filter according to attributes, and store the filtered attribute values in a new dictionary for summarization. The high-frequency words are obtained by unary and binary counts of movie review segmentations, as shown in

$$\text{TF - IDE} = \frac{N_{i,j}}{\sum_k N_{i,j}} \log \frac{D}{1 + |\{j: T_j \in D_j\}|}. \quad (3)$$

Among them, TF is the word frequency, and IDE is the reverse file frequency, which is used to measure the prevalence of occurrence. It is used in the film review of "Shinshaw Redemption," and the high-frequency words and weights are shown in Table 1.

In order to better introduce the above evaluation criteria, first establish a classification about the actual attribution and classifier for a certain text. The recall rate refers to the ratio of the number of texts that actually belong to the category to the number of texts that the classifier discriminates as categories. The calculation formula is

$$r = \frac{a}{a + c}. \quad (4)$$

Accuracy and recall rate reflect the performance of different aspects of the classifier, respectively: accuracy reflects the accuracy of the classifier, and recall rate reflects the integrity of the classifier. The evaluation performance of using accuracy or recall rate as classifier depends on the target concerned by the experimenter. The two evaluation criteria are complementary. Simply increasing one evaluation criterion will lead to a decrease in the other evaluation criterion. A good classifier should have both high precision and recall. In this paper, the precision and recall are synthesized, and the calculation formula is

$$F_\alpha = \frac{(\alpha^2 + 1) \times p \times r}{\alpha^2 \times p + r}. \quad (5)$$

Among them,  $\alpha$  is an adjustable parameter. In text classification,  $\alpha = 0$  is often used, and  $F_1$  is used as the evaluation function. The calculation formula is

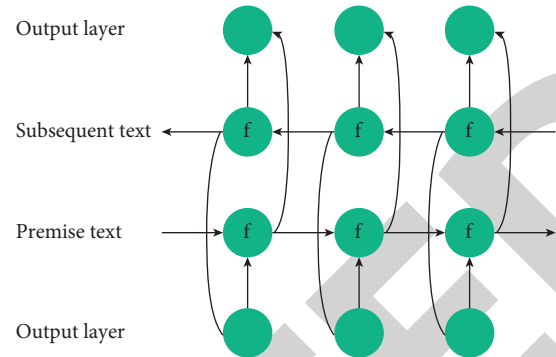


FIGURE 2: Comment text mining.

$$F_1 = \frac{2p \cdot r}{p + r}. \quad (6)$$

Macro-average and micro-average for classification results, it is sometimes necessary to evaluate the global meaning, and all categories of the text collection must be considered. Therefore, two evaluation criteria, macro-average and micro-average, are introduced. Micro-average, also known as micro-average, treats individual texts in a collection of texts as equally important and susceptible to large categories. The calculation formulas of the precision and recall of the micro-average are

$$\text{Micro} = \frac{\sum_{i=1}^C a_i}{\sum_{i=1}^C a_i \sum_{i=1}^C b_i}. \quad (7)$$

Studies based on semantic tendency believe that words in some sentences contain emotional information, and the semantic tendency and intensity of words can be used to determine the semantic tendency of sentences and the entire text. The semantic tendency of a word can be measured from two aspects, direction, and intensity. The direction of a word refers to the positive or negative meaning a word has, and the intensity refers to the degree to which the word is positive or negative. Methods that use semantic orientation, generally do not require training on the training set, are unsupervised learning methods. The steps of this method generally first perform part-of-speech tagging on subjective texts, extract words or phrases containing semantic tendencies, calculate whether these phrases represent positive or negative meanings, that is, calculate semantic tendencies, and finally average all semantic tendencies. Value, the text is positive; otherwise, it is negative. The key steps of the semantic orientation-based method are described below, including word segmentation and part-of-speech tagging, obtaining sentiment words, calculating the semantic orientation of words, and integrating semantic orientations to obtain the semantic orientation of sentences or documents.

## 4. Experimental Analysis and Verification

**4.1. Epoch Parameter Comparison Experiment.** In order to verify the influence of different parameters of the model on the classification effect of the model in this paper, experimental parameters with different Epoch and Batchsize values

TABLE 1: Review high-frequency words and weights.

Output layer	Weight	Key word	Weight	Key word	Weight	Key word	Weight
Key word	0.15	Hero	0.03	Isle	0.08	Money	0.12
Ending	0.10	Moving	0.06	Friendship	0.05	Will	0.06
Marvellous	0.06	Prison	0.13	Law	0.03	Clever	0.08
Plot	0.12	Free	0.17	Friend	0.05	Beer	0.12
Last	0.07	Rebirth	0.05	Study	0.04	Roof	0.11

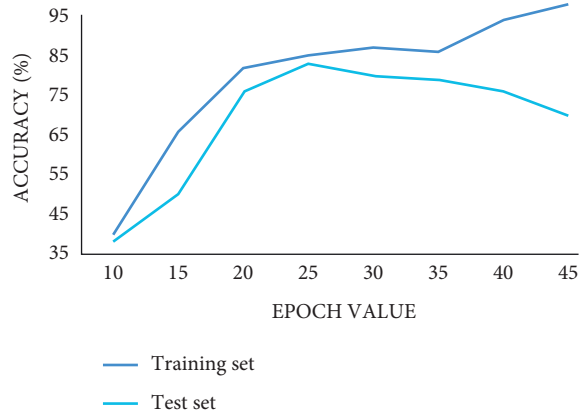


FIGURE 3: Line chart of the accuracy of different Epoch values.

were selected for comparative experiments to explore the influence of parameters on the experimental results. In the learning task of machine learning, Epoch means that a data set has completely passed the review text mining once and returned once. This process is called an Epoch; that is, the review text mining performs a forward pass on all training samples at the same time. One more backpropagation is performed. In the comparative experiment, based on the review text mining model in this paper, different Epoch parameter values are used, and the other model parameters remain unchanged for comparative experiments. The experimental results are shown in Figure 3.

As can be seen from Figure 3, the accuracy of training set and test set usually shows an upward trend with the increase of epoch. At the same time, the accuracy on the test set began to decline, and the reason for the analysis was that the model had begun to overfit. From this, it can be concluded that choosing a suitable Epoch value is very important for machine learning training. When the number of training times is too small, the model will not be fully learned. Rate decreases, resulting in the opposite effect.

**4.2. Batchsize Parameter Comparison Experiment.** Batchsize refers to the number of samples taken in a training process. Batchsize affects the optimization degree and speed of the model. When the value of Batchsize is too small, the training gradient will be too large, making it difficult for the model to converge. When the value of Batchsize is too large, although there is a suitable gradient to speed up the training of the model, it will greatly increase the amount of calculation, so that the training time is too long and resources are wasted. In order to study the influence of the Batchsize parameter on the experimental effect, a comparative

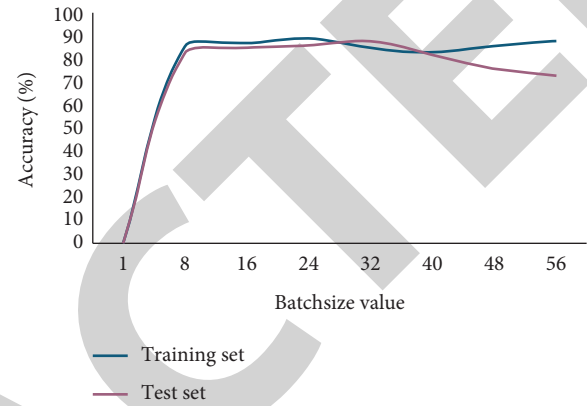


FIGURE 4: Accuracy model for different Batchsize values.

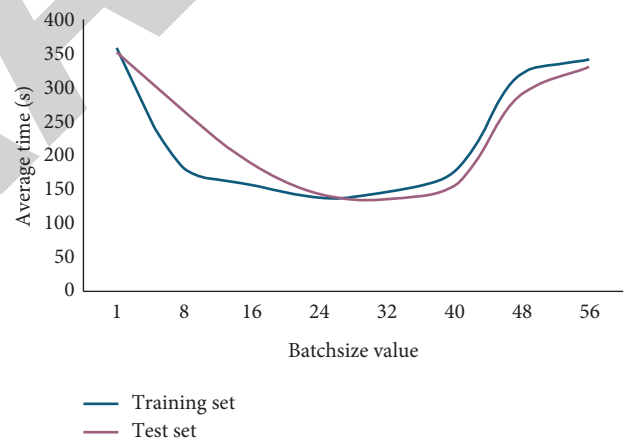


FIGURE 5: Average consumption time for different Batchsize values.

experiment was carried out under the condition of setting different Batchsize values and keeping other model parameters the same. The classification effect and average time of the obtained comparative experiments are shown in Figures 4 and 5.

As can be seen from Figures 4 and 5, when the Batchsize is small, the accuracy of the model is not improved, and the time spent in each iteration process will also be too large, which will eventually slow down the training of the model. When the Batchsize is very large, each iteration consumes more and more time, and the training process of the model wastes too much memory. A suitable batch not only takes a short training time, but also takes up less time resources. In the comparative experiment, when the Batchsize is 32, from the perspective of time overhead, memory consumption, and accuracy, higher model accuracy can be obtained, while

TABLE 2: Experimental results of text sentiment classification.

Models		UE	HE	IE	NE
100d	Precision	92.26%	90.66%	91.26%	86.91%
	Recall	91.05%	89.80%	90.62%	87.49%
	F1	91.67%	90.24%	90.96%	87.21%
150d	Precision	92.76%	91.41%	91.55%	87.71%
	Recall	91.45%	90.44%	91.92%	87.88%
	F1	92.12%	91.43%	91.24%	87.77%
200d	Precision	91.46%	90.91%	91.05%	86.71%
	Recall	91.26%	90.04%	90.42%	87.33%
	F1	91.85%	90.91%	90.72%	86.98%

the time consumption is less and the convergence speed is faster. Through comparative experiments, it can be concluded that when Batchsize = 32 and Epoch = 25, the model achieves the optimal classification effect.

**4.3. Experimental Results of Sentiment Classification of Film and Television Texts.** The sentiment analysis evaluation indicators in this experiment use several indicators commonly used in the field of sentiment analysis: precision, recall, and F1 value. TP represents the number correctly classified as positive, FP represents the number incorrectly classified as positive, TN represents the number correctly classified as negative, and FN represents the number incorrectly classified as negative.

The accuracy rate represents how many of the test samples judged as positive by the model are positive, and the recall rate represents how many of the samples judged by the model to be the same as the actual samples are positive samples. According to the precision and recall, the classification performance of a classifier for positive samples and negative samples can be objectively evaluated.

In the experiment, 200-dimensional, 150-dimensional, and 100-dimensional word vectors were used for training on the four models. The experimental data are shown in Table 2.

From Table 2, the following conclusions can be drawn.

When the dimension of the word vector is fixed, the number of convergence rounds and the average training time of the NE model are both the highest. In contrast, the effect of IE on these two indicators is better than that of HE. UE combines the results of IE and HE. The advantage is that the number of convergence rounds and the training time are relatively low.

Figures 6 and 7 show the impact of the experimental results on the model from the perspective of word vector dimension changes.

The above figure is a schematic diagram of the changes of various indicators when the joint embedding topic word vector model UE adopts 200-dimensional, 150-dimensional, and 100-dimensional word vectors. When other models use word vectors of different dimensions, the trend of changes in indicators is similar to that of UE, so horizontal comparison of various indicators is the conclusion that UE has the best effect.

Use different classification algorithms to build sentiment classifiers in the training data, use the test data to determine the accuracy of the classifier, use the classifier to predict the accuracy, and then use the accuracy to select the best

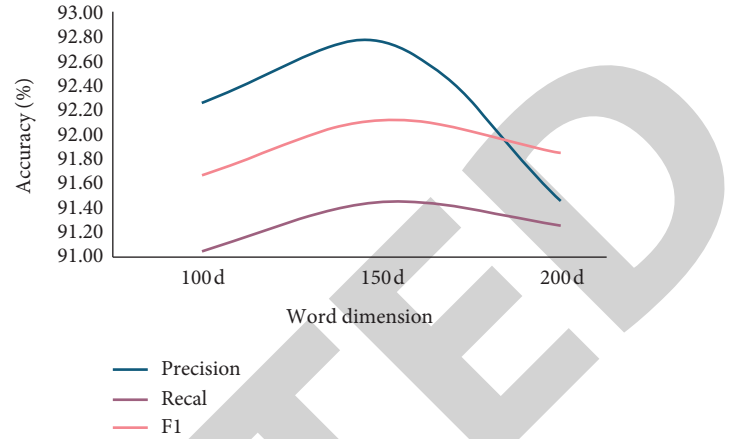


FIGURE 6: Changes in UE precision, recall, and F1 value.

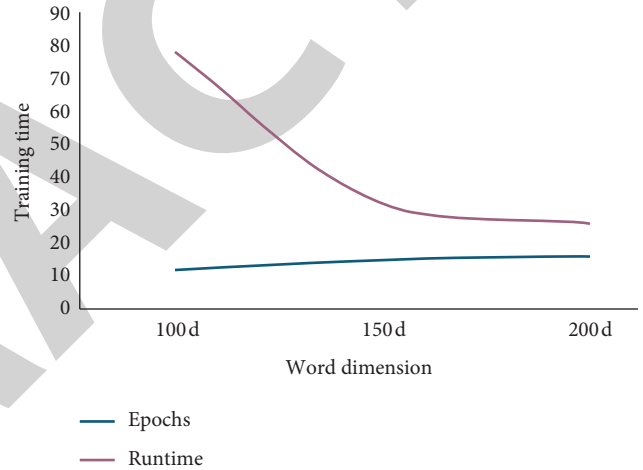


FIGURE 7: Changes in the number of UE training rounds and training duration.

classification algorithm. Adjust the number of features to test the accuracy of the classifier to make it more accurate. The sentiment score algorithm constructed by two algorithms, Naive Bayes and Support Vector Machine, can capture the dynamics of the real scene of the graph sequence data type.

## 5. Conclusions

In this paper, text mining technology and machine learning algorithm are applied to the identification and analysis of movie elements, and through descriptive analysis and sentiment analysis, and we try to find the useful information hidden behind the identification of user elements. In the descriptive analysis, the method of calculating the probability of high-frequency words and generating the graph cloud is used to express it intuitively and visually. At the sentiment analysis level, the element recognition text is vectorized, classified by the naive Bayesian classification method, the high probability subject words are found, the advantages and disadvantages of the specified movies are



analyzed, and the deep subject mining is carried out. Under the background of big data, the in-depth analysis of the element identification of moviegoers is conducive to the objective and overall evaluation of the movie, to reflect the true feelings of moviegoers, and to help other moviegoers to make decisions on whether to watch the movie, which has certain practical significance. The creation of film and television dramas is an extremely complex proposition. The market share and success of the works require not only quantitative analysis and evaluation of big data, but also the rational use of many uncontrollable factors and quantifiable elements to ensure the birth of an excellent work. Data analysis can never replace content production. Therefore, with the ever-changing technological changes and thinking changes, it is necessary to give full play to the strong scientific data support provided by big data for the planning, creation, and publicity of new media original film and television dramas. Good at analyzing data, judging audience needs, understanding audience behavior, and increasing the success rate of original film and television dramas, especially high-quality drama projects, only a chemical reaction of various elements can produce an excellent original film and television work that can stand the test of the audience. However, there are still some problems in this paper. For example, data analysis can never replace content production. The simulation verification of the model needs further elaboration and analysis in the future research.

## Data Availability

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

## Conflicts of Interests

The author declares no conflicts of interest.

## References

- [1] B. Liu, Z. Lv, N. Zhu, and D. Chang, "Research on the evaluation of the dissemination ability of sci-tech periodicals based on hesitant fuzzy linguistic," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 28, no. 2, pp. 153–167, 2020.
- [2] N. Fang, "Psychological influence of Chinese film and television works in teaching Chinese as a foreign language," *International Journal of Electrical Engineering Education*, Article ID 002072092110020, 2021.
- [3] Z. Takeoka, *Text Mining Using Machine Learning (2) Discovery and Verification of Hypotheses*, p. 60, Collection of Essays on Momoyama Gakuin University, Osaka, Japan, 2019.
- [4] B. E. Howard, J. Phillips, K. Miller, and A. Tandon, "SWIFT Review: a text mining workbench for systematic review," *Systematic Reviews*, vol. 5, no. 1, 2016.
- [5] K. Christensen, S. Nørskov, L. Frederiksen, and J. Scholderer, "In search of new product ideas: identifying ideas in online communities by machine learning and text mining," *Creativity and Innovation Management*, vol. 26, no. 1, pp. 17–30, 2016.
- [6] W. U. Songzhi, "On translation skills of cultural elements in subtitles of Japanese film and television," *Journal of the Chinese People's Armed Police Force Academy*, vol. 15, no. 1, pp. 59–68, 2015.
- [7] X. Xiao and X. Zhao, "Evaluation of mangrove wetland potential based on convolutional neural network and development of film and television cultural creative industry," *Arabian Journal of Geosciences*, vol. 14, no. 12, 2021.
- [8] Q. I. Lisen, "Research on the interactive mechanism between creative teaching of film and television advertising and college students competition," *The Theory and Practice of Innovation and Entrepreneurship*, vol. 31, no. 20, p. 12, 2019.
- [9] S. V. Ravi, B. D. Kumar, S. Raj, and A. Divyacharan, "Text mining methods for online topics and reviews using machine learning," *Journal of Physics: Conference Series*, vol. 1916, no. 1, Article ID 012215, 2021.
- [10] J. Scholderer, L. Frederiksen, K. Christensen, and N. Sladjana, "In Search of new product ideas: identifying ideas in online communities by machine learning and text mining," *Creativity and Innovation Management*, vol. 26, no. 1, pp. 17–30, 2017.
- [11] Q. Zhu, Y. Wu, Y. Li, J. Han, and X. Zhou, "Text mining based theme logic structure identification: application in library journals," *Library Hi Tech*, vol. 36, no. 3, pp. 411–425, 2018.
- [12] M. Short, "Text mining and subject analysis for fiction; or, using machine learning and information extraction to assign subject headings to dime novels," *Cataloging & Classification Quarterly*, vol. 57, no. 5, pp. 1–22, 2019.
- [13] M. Bruynooghe, H. Blockeel, B. Bogaerts et al., "Predicate logic as a modeling language: modeling and solving some machine learning and data mining problems withIDP3," *Theory and Practice of Logic Programming*, vol. 15, no. 6, pp. 783–817, 2015.
- [14] R. Samizade and E. M. S. Abad, "The application of machine learning algorithms for text mining based on sentiment analysis approach," *Journal of Information Technology Management*, vol. 7, no. 02, pp. 10–25, 2018.
- [15] Z. Li and A. Cao, "A creative study on wushu heroism in animation film and television-take the animation kung fu panda as an example," *Chinese Wushu Research*, vol. 5, no. 32, pp. 96–112, 2017.
- [16] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current state of text sentiment analysis from opinion to emotion mining," *ACM Computing Surveys*, vol. 50, no. 2, pp. 1–33, 2017.
- [17] R. Deotale, S. Rawat, and V. Vijayarajan, "POCASUM: policy categorizer and summarizer based on text mining and machine learning," *Soft Computing*, vol. 25, no. 14, pp. 9365–9375, 2021.
- [18] E. Popoff, M. Besada, J. P. Jansen, S. Cope, and S. Kanters, "Aligning text mining and machine learning algorithms with best practices for study selection in systematic literature reviews," *Systematic Reviews*, vol. 9, no. 1, p. 293, 2020.
- [19] Y. Huang, Y. L. Murphey, and Y. Ge, "Intelligent typo correction for text mining through machine learning," *International Journal of Knowledge Engineering and Data Mining*, vol. 3, no. 2, p. 115, 2015.
- [20] G. Kaur and M. Kumar, "A review on sentiment analysis of social media data using text mining and machine learning," *International Journal of Advanced Research*, vol. 4, no. 5, pp. 772–775, 2016.
- [21] X. Xiao, Y. Cheng, and J. M. Kim, "Movie title keywords: a text mining and exploratory factor Analysis of popular movies in the United States and China," *Journal of Risk and Financial Management*, vol. 14, no. 2, p. 68, 2021.
- [22] Y. Taniguchi and K. Tsuda, "A study of the possibilities of text mining and machine learning for score evaluation and review

content,” *Artificial Intelligence Research*, vol. 9, no. 1, p. 1, 2020.

- [23] S. Chowdhury and S. Gupta, “A comparative study of text mining in big data analytics using deep learning and other machine learning algorithms,” *International Journal of Hybrid Intelligence*, vol. 1, no. 2/3, p. 163, 2019.
- [24] J. Saarikoski, H. Joutsijoki, K. Jarvelin, and M. Juhola, “On the influence of training data quality on text document classification using machine learning methods,” *International Journal of Knowledge Engineering and Data Mining*, vol. 3, no. 2, pp. 143–169, 2015.

RETRACTED