International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-13

https://doi.org/10.70917/ijcisim-2025-0246
Article

Analyzing the Characteristics of Female Emotional Expressions in the Works of Chinese Filipino Women Writers Based on Emotional Computing Models

Ting Shao [1,2,*]

[1] De La Salle University, Manila, 0922, Philippines
[2] Shangqiu Normal University, Shangqiu, Henan, 476000, China; tingtingxx@126.com

Abstract: Applying sentiment recognition and computation to literary works has gradually become a popular direction in recent digital humanities research. The study introduces the sentiment computation model into the analysis of Filipino Chinese female writers' works, effectively extends the basic sentiment lexicon through the improved SO-PMI algorithm, constructs the sentiment computation model, and carries out experimental analyses and empirical studies on the sentiment analysis of Filipino Chinese female writers' works. The experimental analysis verifies the advantages of the proposed sentiment computation model in the task of text sentiment analysis as well as the effectiveness of the sentiment lexicon, and the F1 values of the improved SO-PMI algorithm for the classification of positive, neutral, and negative sentiments are improved by 11.96%, 6.96%, and 6.89% over the SO-PMI algorithm. The percentage of emotions in the works of Filipino Chinese women writers is dominated by sadness, fear, and anger, which all account for more than 16%, reflecting the difficult situation, identity entanglement and inner sorrow of Filipino expatriates. The emotion calculation model can reveal and summarize the structure of emotion in narrative texts, providing a richer and more three-dimensional perspective for Chinese literary criticism.

Keywords: SO-PMI algorithm; affective computational model; affective lexicon; Chinese literature

1. Introduction

As an ethnic minority in the mainstream Philippine society, the literary works of Chinese Filipino women writers used to be ignored and marginalized for a long time [1-2]. It is only after several generations of efforts that Chinese literature has begun to receive attention from the mainstream Filipino society in recent decades, showing increasing prosperity [3-4]. In the process of the development of Filipino Chinese literature from neglect to attention, a large number of Chinese women writers and their works have played a very important role. Reading their works carefully, it is not difficult to find that most of these women writers have expressed their search and reflection on the construction of their own cultural and gender identities, as well as their struggle against gender and racial discrimination through the narration of their own and their parents' generation's life experiences, and the presentation and exploration of their ethnic histories and cultures from a uniquely female point of view [5-9]. Therefore, feminism is an important theme in their works [10].

In the world of literature, Chinese women writers have depicted the picture of life with unique perspectives and delicate strokes, creating a colorful image [11-12]. Some of these images are as tough as stone, some are as gentle as water, and they are either marching bravely in the torrent of the times, or wandering and struggling in the entanglements of emotions [13-14]. In literature, the works of female writers can be very different from those of male writers. Female writers have unique styles and characteristics in their literary creation and expression, and their works pay more attention to communication with readers, focusing on the expression of human reality, focusing on women's issues, as well as emphasizing the beautiful details of life [15-19]. The works of female writers also use social realities and reflections, as well as more creative elements, which make their works better literary chapters and make them popular and respected in the field of contemporary literary creation [20-22].

The study explores the expression of women's emotions in the works of Filipino-Chinese women writers, and summarizes the characteristics of their expressions in terms of the expression of vernacular sentiment, the criticism of traditional concepts and cross-cultural conflicts. On this basis, the basic emotion lexicon, the implicit emotion lexicon, the degree adverbial lexicon and the negation lexicon were collected and organized to construct the emotion lexicon. The TF-IDF is used to select the benchmark words, and the classifier polynomial Bayesian classifier is used to realize the extraction of candidate words, and then the remaining words are selected by the improved SO-PMI algorithm to find out the most emotionally representative words, to complete the classification of emotional tendency of the text. Subsequently, the experimental analysis of the sentiment computing model and the sentiment analysis of the works of Filipino Chinese female writers are conducted. This paper introduces the sentiment calculation method into the field of literature, and quantitatively analyzes the sentiment of the texts of Filipino Chinese women writers' works, which fills a key missing piece in the study of metric literature, and also provides new ideas and methods for the study of Chinese literature.

2. Expression of Female Emotions in the Works of Filipino-Chinese Women Writers

Filipino Chinese literature shares the same language and origin with Chinese literature, and is an important part of the world's Chinese literature. This paper selects the novels written by Filipino-Chinese women writers as the objects to explore the characteristics of female emotional expression in them. The works of Filipino-Chinese women writers focus on the discovery of the spirit of the national culture, and they raise a number of issues from traditional Chinese culture that have been neglected in the past with a woman's perspective, which includes the promotion of the ideal character of women that embodies the excellent traditions of the nation, as well as the diagnosis and critique of the national spirit of the suppression of women's illnesses.

2.1. Expression of rustic sentiment

The works of Filipino Chinese women writers are often characterized by their attachment to their country of origin, ethnicity, and kinship, showing a distinctive blend of Chinese narratives and local consciousness. Xiao Si's The Lilac Knot shows the complexity of nostalgia and sentimentality on a middle-aged man who is nostalgic for his homeland. For the protagonist Pan, nostalgia and sentimentality have the dimensions of real homeland and homeland people, as well as the dimensions of cultural significance. The nostalgia and sadness on the level of the real homeland and people become the framework of the story, while the cultural dimension is embedded in the meaning and background of the story itself. Most of the works of Filipino-Chinese women writers are filled with a kind of attachment to the nation and culture of their country of origin, and the promotion of the significance of the national culture itself.

2.2. Traditional Concepts and Value Criticism

The creative vision of Filipino-Chinese women writers has been quite focused on the plight of women in the diaspora, the phenomenon of the “Auntie Pancake.” This type of storytelling depicts women who are persistent, determined, tolerant, and self-sacrificing in the face of hardship, and portrays the image of the sacrificial wife as a stone of hope in the Chinese cultural context. At the same time, it denies and rejects women's self-sacrifice, which is full of alienating colors, but does not lead to a good ending. Xiaohua's Youth in the Ashes reflects on the traditional view of women from different perspectives, and writes about how the traditional concepts of “a woman should be married when she is old enough” and “a widow should keep her promise” have become an invisible pressure, and how the traditional concepts of women's rights have become an invisible pressure, and how they have become an invisible pressure. The author reflects on the traditional view of women from different perspectives, writing about how traditional concepts such as “women should be married” and “widows should keep their holidays” have become an invisible pressure on two young women with different destinies, which implies the author's rejection of the traditional irrational moral principles. However, the writer pours his sympathy on this kind of persistent, determined, tolerant and self-sacrificing women, artistically records their voices of the heart and their journey, and reveals the truth of women's existence, which has been concealed and suppressed by history and life.

2.3. *Cross-cultural conflict and integration*

When examining and investigating cross-cultural love and marriage phenomena, Filipino-Chinese female writers place the conflicts and contradictions in gender relations within the context of the collision and integration of diverse cultures, thereby contemplating issues such as "gender", "marriage", "ethnicity", and "culture" in a broader tragic context. The marital conflict caused by the clash of two heterogeneous cultures is a reality of particular concern to women writers. Chen Qionghua's “Friends and Children” is about a family with two cultures mixed together: the husband, Chen Zhongjun, is Chinese, and the wife, Yaling Yi, is Filipino; there is obviously a clash of two different ethnic cultures in the education of their children and in the etiquette of family and friends' interactions. Pei Qiong's “Oil Paper Umbrella” is also a love tragedy caused by the ethnic and cultural barriers. The heroine, Jenny, is a half-Chinese, half-Filipino girl, who is in love with her boyfriend, but is opposed by the boyfriend's mother. The author uses Jenny's repeated recitation of the poem “Oil Paper Umbrella” and her preference for “oil paper umbrellas”, which is full of Chinese folklore, to write about her identification with and love of Chinese literature and culture, lamenting that her own “Chineseization” cannot change the secular racial prejudice, and criticizing the narrow-mindedness of traditional culture. She also criticizes the narrow-mindedness of traditional culture.

3. Improved SO-PMI-based model for sentiment calculation

In order to be able to access the emotions in the works of Filipino-Chinese female writers and analyze the characteristics of their emotional expressions, this chapter constructs an emotional computation model based on the improved SO-PMI, which includes text preprocessing, emotion lexicon construction, and the improvement of the SO-PMI algorithm.

3.1. *Text Sentiment Analysis Process*

Text Sentiment Analysis, also called Opinion Mining, refers to the analysis and processing of text with emotional color to dig out the emotions contained in the text, and analyze and classify the obtained emotions. As a research direction of natural language processing, text sentiment analysis has the process of opinion analysis, processing, induction and reasoning. The process of text sentiment analysis is shown in Figure 1. Data preprocessing refers to the deactivation of data and word separation, and common methods include removing invalid characters and data, using word separation tools for word separation processing, and deactivation word filtering. After word separation, the text is extracted from the sentiment words, and the sentiment words are calculated to finally get the sentiment value of the text, and the text is categorized according to the sentiment value.

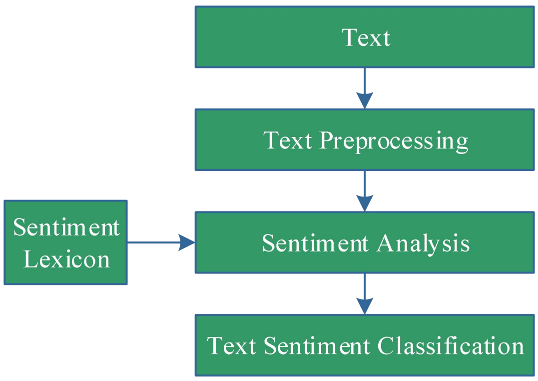

Figure 1. Text emotional analysis process.

3.2. *Text pre-processing*

The original text obtained through the information acquisition technology can not be directly used for information processing, the text must be converted into structured data for computer recognition through text pre-processing, i.e., the text is formalized. Text pre-processing technology mainly includes word separation technology and removal of stop words, text representation and feature extraction. Segmentation is the use of segmentation tools to divide the text into shorter words. Removal of deactivated words, is in accordance with a certain method, to remove the text generally does not include effective text nature of the pronouns, prepositions, auxiliaries and other function words. In this paper, the jieba disambiguation technique was chosen to analyze the text of the works of Filipino Chinese women writers. The skipgram model in word2vec is used to train the word vectors.

3.3. Emotional lexicon construction

The existing sentiment dictionaries are all applicable only to comment texts. Therefore, it is necessary to optimize and expand the sentiment lexicon. In this paper, we choose the sentiment vocabulary ontology as the general sentiment dictionary, add implicit sentiment words on its basis, and then expand it using the SO-PMI algorithm.

3.3.1. Basic Affective Dictionary

Currently existing Chinese emotion dictionaries only categorize emotion words into positive emotion words, negative emotion words, positive words, negative words, and neutral words, and for the textual emotion classification of the works of Filipino-Chinese female writers, it is not enough to obtain the emotion polarity of the words. The Sentiment Lexical Ontology is a Chinese lexical ontology resource with sentiment weights labeled by the Information Retrieval Department of Dalian University of Technology. This resource describes words or phrases from different perspectives, mainly including the sentiment category, lexical category, sentiment intensity and polarity of words. Therefore, the sentiment lexical ontology is chosen as the basic sentiment lexicon.

3.3.2. Implicit Sentiment Dictionary

Most of the works of Filipino-Chinese women writers are fiction texts, a literary genre that generally centers on characters and depicts social life through the setting in which they live and the stories they tell. Character, plot, and setting are the three elements of a novel. There are many writing techniques and ways of depicting emotions in the works of Filipino-Chinese women writers, and it is not enough to analyze the text using only the vocabulary of emotions. The more techniques used to describe the novel, the more difficult it is to extract the emotional vocabulary of the text. Through the study of novel writing techniques, this paper chooses more suitable symbolic words to build an implicit emotion lexicon. Symbolism is the use of certain things to convey special meanings. For instance, "bright moon" is used to convey "yearning", and "willow tree" represents "parting", etc. These words have no obvious emotional tendency, but they imply certain emotions. In this paper, the symbolic words are extracted and the basic sentiment lexicon is expanded, so as to enlarge the number of words in the lexicon and improve the accuracy of sentiment analysis of long texts. The emotion lexicon ontology is used as the basis to assign emotion to the symbolic words. A total of 100 symbolic words are collated, of which 65 are positive implicit sentiment words and 35 are negative implicit sentiment words.

3.3.3. Sentiment Dictionary of Adverbs of Degree

In Chinese emotional expression, adverbs of degree are often used to modify emotional words. Although the emotional polarity of a text is mainly determined by the words used to express emotion, the degree of its strength is mainly reflected by the degree adverbs, which are usually used in front of the modified emotion words to change the strength of the overall emotional expression of the sentence. Based on the adverbs of degree in the Knowledge Network Emotion Dictionary, this paper selects a total of 50 commonly used adverbs of degree.

3.3.4. Dictionary of Negative Emotions

In the expression of the works of Chinese Filipino women writers, negatives have a direct effect on the emotional tendency of the text. When there is no negative word in front of it, the tendency of the emotional word can directly affect the emotional tendency of the text. When there is a negative word in front of it, the emotional tendency will be opposite to the emotional tendency of the emotional word. Therefore, when performing the emotional calculation of the works of Filipino Chinese women writers, it is necessary to consider whether the negative words have an effect on the emotional tendency of the text, and there are 54 negative words collected and organized in this paper.

3.4. Calculation of affective tendencies

In order to be able to expand the sentiment words as much as possible, this paper uses the SO-PMI algorithm to obtain the sentiment words that have not been added to the sentiment dictionary. In this paper, the improved SO-PMI algorithm is used to complete the automatic construction of the sentiment lexicon to better perform the task of text sentiment analysis.

3.4.1. TF-IDF based base word selection

In practice, there are two kinds of words, one is the word frequency is very low, only once or a few times, but they all appear in the same emotional tendency of the text, using the corresponding algorithm to calculate them but have a strong emotional tendency, for example, a word only once in the negative text occurs, directly when using the SO-PMI algorithm, it will be obtained when the emotional tendency of the value of the value of the very large, but it does not have a representative. Another kind of words are those that appear very frequently in both positive and negative texts, but they do not have practical meaning, such as the in English, so the TF-IDF algorithm is used to filter out those low-frequency words and words with high frequency but no practical meaning:

$$TF = \frac{W_n}{W_{all}} \tag{1}$$

$$IDF = \log \frac{D_{all}}{D_n + 1} \tag{2}$$

$$TF - IDF = TF * IDF \tag{3}$$

In Equation (1) W_n is the word frequency value of a single word and W_{all} is the sum of word frequencies of all the words. This reflects how much the word frequency of a single word is in the sample data.

D_{all} is the number of all data texts, D_n is the number of texts including the word n, and 1 is added in the formula in order to avoid the error of using the TF-IDF algorithm with a denominator of zero. The final value of TF-IDF is obtained by multiplying the value of TF and the value of IDF.

After obtaining the value of TF-IDF, a certain threshold can be set, and when the value of the requested TF-IDF is less than the threshold, it is discarded. When the value is greater than the threshold, the next step of solving for the SO-PMI value can be performed. This reduces the influence of some low-frequency extreme words and words with insignificant emotional tendencies on the construction of the whole sentiment dictionary, and also reduces a certain amount of workload.

3.4.2. MNB-based Candidate Word Extraction

On this basis, a candidate word extraction method based on the classifier Multinomial Bayesian Classifier (MNB) is proposed, which can be expanded by generating an extended sentiment lexicon using the improved SO-PMI algorithm after optimizing the selection methods of the benchmark and candidate words.

Multinomial Plain Bayes (MNB) views a document as a bag-of-words model and considers that the frequency of occurrence of words in a document has an impact on the prediction of document categories. Therefore, when calculating conditional probabilities, MNB needs to count the frequency of word occurrences.

In the MNB model, a text can be represented as a vector $w_i \in N$, $w_i \in N$, where w_i denotes the frequency of occurrence of a word in document d. Like Bernoulli's plain Bayesian model, MNB makes a conditional independence assumption so that different conditional probability estimates do not affect each other. Given the document d to be tested, the MNB model predicts the document d using the following formula:

$$c(d) = \arg\max_{c \in C} \left[\log_2 p(c) + \sum_{i=1}^{m} f_i \log_2 p(w_i \mid c) \right] \tag{4}$$

where the a priori probability $P(c)$ can be computed by Equation (5), which also employs Laplace estimation, and l denotes the number of attribute values for the category attribute C. Then:

$$p(c) = \frac{\sum_{j=1}^{n} \delta\left(c_j, c\right) + 1}{n + l} \tag{5}$$

The conditional probability $P(w_i \mid c)$ denotes the probability of occurrence of word w_i in a sample belonging to class c, and the conditional probability $P(w_i \mid c)$ is computed using equation (6):

$$p(w_i \mid c) = \frac{\sum_{j=1}^{n} f_{ji}\delta(c_j, c) + 1}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ji}\delta(c_j, c) + m} \tag{6}$$

where f_{ji} denotes the frequency of the i th word in the j th document, c_j denotes the class labeling of the j th document, and m denotes the number of attributes.

3.4.3. Improved SO-PMI algorithm

The original PMI algorithm is shown in equation (7):

$$PMI(w_1, w_2) = \log_2 \frac{P(w_1, w_2)}{P(w_1)P(w_2)} \tag{7}$$

The original PMI algorithm is shown in equation (7):

$$\begin{aligned} SO-PMI\left(word\right) = &\sum_{Pword \in Pwords} PMI\left(word, Pword\right) \\ &- \sum_{Nword \in Nwords} PMI\left(word, Nword\right) \end{aligned} \tag{8}$$

SO-PMI calculates the correlation between two words, and when calculating the SO-PMI value, the negative emotion seed words and positive emotion seed words are divided first, and then the words to be sought and many seed words are calculated separately, and then subtracted to find the emotional tendency of the word. Such an emotion dictionary construction method has great limitations. Therefore, some adjustments are made to the SO-PMI algorithm, instead of using the correlation between words and seed words, the correlation between words and the corresponding emotional tendency text is used to calculate the SO-PMI value of the words, and the value is used to automatically complete the automatic construction of the emotion dictionary, firstly, the two formulas for PMI are given:

$$PMI(w, pos) = \log \frac{p(pos \mid w)}{p(pos)} \tag{9}$$

$$PMI(w, neg) = \log \frac{p(neg \mid w)}{p(neg)} \tag{10}$$

Where w denotes any word in the text, in this case it is the word that has been filtered by the TF-IDF algorithm in our experiment. $p(pos)$ is the probability of labeling as positive sentiment tendency in the whole text training set, and $p\left(pos \mid w\right)$ is the probability of occurrence of sentences labeled as positive in all the sentences containing w in the text training set. When $PMI\left(w, pos\right)$ means the relevance of the positive sentiment of the word w, when $PMI\left(w, pos\right)$ is greater than 0, the larger its value is, the stronger the positive sentiment of the changed word. When $PMI\left(w, pos\right)$ is greater than 0, the larger the value is, the weaker the positive sentiment of the word w. The opposite is true for $PMI\left(w, neg\right)$.

After obtaining $PMI\left(w, neg\right)$ and $PMI\left(w, pos\right)$, according to the original SO-PMI algorithm, we can calculate the affective tendency of the word w by subtracting the two values $SO\left(w\right)$:

$$SO\left(w\right) = PMI\left(w, pos\right) - PMI\left(w, neg\right) \tag{11}$$

When $SO\left(w\right)$ is greater than 0, the sentiment polarity of word w is positive and the larger the value, the stronger the positive sentiment, when the value of $SO\left(w\right)$ is close to 0, the word w is close to a neutral word, and when the $SO\left(w\right)$ of word w is smaller than 0, it represents that the sentiment tendency of word w is negative and the smaller the value, the stronger the negative sentiment.

The difference between the maximum and minimum values in the obtained results is very much, so the normalization work should be carried out to facilitate the subsequent work.

In this experiment, linear normalization, which is min-max normalization, is used, and only relevant linear transformations are performed on the obtained SO-PMI values without changing their properties.

So that the final values obtained can be mapped between the interval $[0,1]$. The specific transformation function is as follows:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{12}$$

Where: x is to be normalized, $\min(x)$ is the minimum of all values, $\max(x)$ is the maximum of all values, and x' is the result obtained after normalizing the values.

Thus, in order to reflect the word frequency of the word into the equation, the open square of the word frequency of the corresponding word w in the corpus is added to the whole $SO(w)$ equation to get the new $SO(w)$ calculation equation. That is:

$$SO(w) = \left[PMI(w, pos) - PMI(w, neg)\right] * \sqrt{f(w)} \tag{13}$$

where $f(w)$ is the word frequency. By doing the above, we then obtain the value of SO-PMI with the word frequency of the word w, which in turn gives us a completely new sentiment lexicon different from the one above.

3.4.4. Constructing sentiment vector words

In the previous work, this paper utilized the skip-gram model in the word2vec model for text vectorization. In order to make up for the insufficiency of the word vectors trained by the word2vec model in the task of sentiment analysis, the values of SO-PMI of the words in the sentiment lexicon are added to the trained word vectors. This is done as follows: first generate a full 1 matrix with the same dimension as the trained word vectors, and then multiply the value of SO-PMI of the word to a new matrix $W_{so\text{-}pmi}$, and then pick the word vectors of the word, i.e. W_{word}. These two matrices are then added together to obtain the desired sentiment word vector:

$$W_{new} = W_{word} + W_{so\text{-}pmi} \tag{14}$$

If the selected negative adverb occurs in the text, multiply the W_{word} of the one word following the negative adverb by -1 and discard the word vector of the degree adverb.

4. Empirical study of affective computing models

4.1. Data set and empirical subjects

This paper adopts the standard microblogging dataset of Task 4 in the Sixth Chinese Propensity Assessment Conference, the size of the dataset is about 40,000 articles, of which the labeled samples are about 7,000 and the interference samples are 33,000, from which 12,000 pieces of data are extracted as the test corpus of the emotion calculation model. Meanwhile, this paper takes the texts of Filipino Chinese female writers' works as the empirical objects, and selects 15 of these literary works for text pre-processing, which will be used as the corpus of this paper to study the characteristics of female emotion expression in the works of Filipino Chinese female writers.

4.2. Experimental analysis of the model

4.2.1. Dictionary comparison experiment

A single base dictionary does not cover most of the words, and the basic words are not very effective in classifying the comment texts in specialized fields, so a comparison experiment between a single base sentiment dictionary and a combined sentiment dictionary was set up. In this experiment, 3,000 pieces of positive texts and 3,000 pieces of negative texts were randomly selected from the dataset.

Based on the merged sentiment lexicon and other lexicons constructed in Chapter 3, the sentiment analysis was performed using the sentiment computing algorithm above. The comparison dictionaries include: HowNet-based sentiment dictionary, NTUSD sentiment dictionary, and Tsinghua University's Li Jun Chinese Positive and Negative Sentiment Dictionary (denoted as PDED), and the results of the different dictionaries' sentiment polarity classification are shown in Figure 2.

The accuracy and recall of the merged-based sentiment dictionary have been improved, and its results in positive text are 75.31% and 70.14%, and in negative text are 73.55% and 67.57%, and its F1 value is the highest for the other three sentiment dictionaries, which are 72.63% and 70.43%, respectively. The HowNet sentiment dictionary (HowNet) has higher accuracy and recall in sentiment analysis compared to the other two sentiment dictionaries, but the separate sentiment dictionaries still have lower accuracy and recall than the combined sentiment dictionary when analyzing online comments for sentiment tendency.

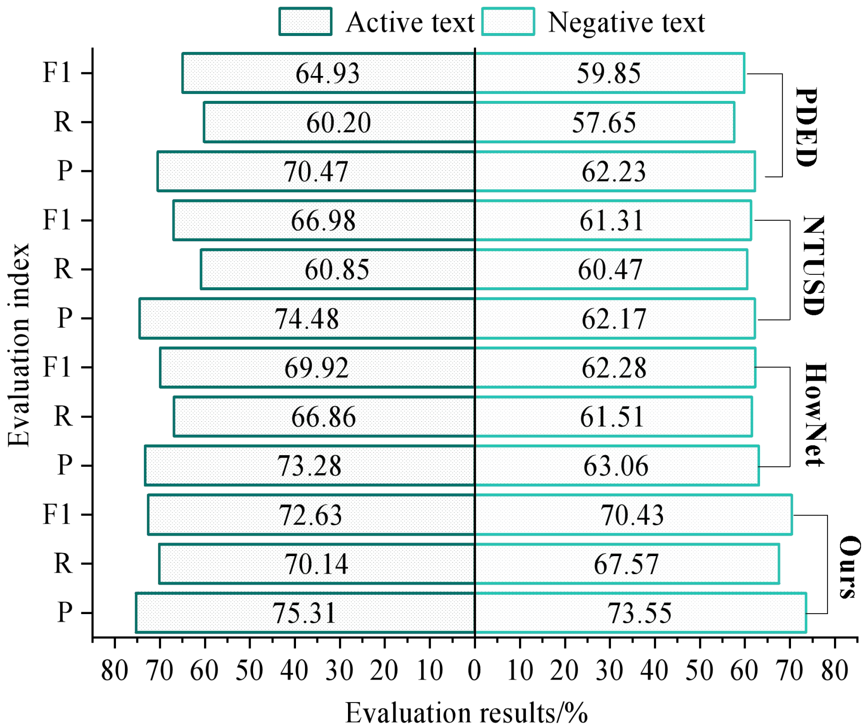

Figure 2. The emotional polarity of different dictionaries.

4.2.2. Analysis of feature selection methods

3000 texts are randomly selected from the dataset as the training set, 2000 texts of positive texts and 1500 texts of negative texts as the test set. Four feature selection methods are defined, namely: 1) words as candidate features, 2) double words as candidate features, 3) words and double words together as candidate features, and 4) Jieba participle processed words as candidate features. Then the top n informative ones are selected as features using chi-square statistics respectively.

The highest accuracy is selected from six common machine learning methods, including Bayesian, logistic regression, support vector machine, etc., which are calculated according to different feature selection methods, and the accuracy of different classification methods is shown in Figure 3. Where BNB is Bernoulli Bayes, MNB is Multinomial Bayes, LR is Logistic Regression, SVC is Support Vector Machine, LinearSVC is Linear Support Vector Machine and NuSVC is Nu Support Vector Machine.

The following two conclusions can be clearly obtained:

(1) Using Jieba participles to select informative words as classification features, the accuracy of all classifications is the highest in this category, with accuracies of 92.72%, 94.88%, 93.39%, 91.36%, 92.25%, 92.76%, which are all above 91%, so Jieba participles should be chosen to select features.

(2) In the case of selecting words and double words as features, and in the case of Jieba participle selecting informative words as features, MNB, LR and LinearSVC have the best classification performance, so these three classifiers should be selected as candidate classifiers for testing classification dimensions.

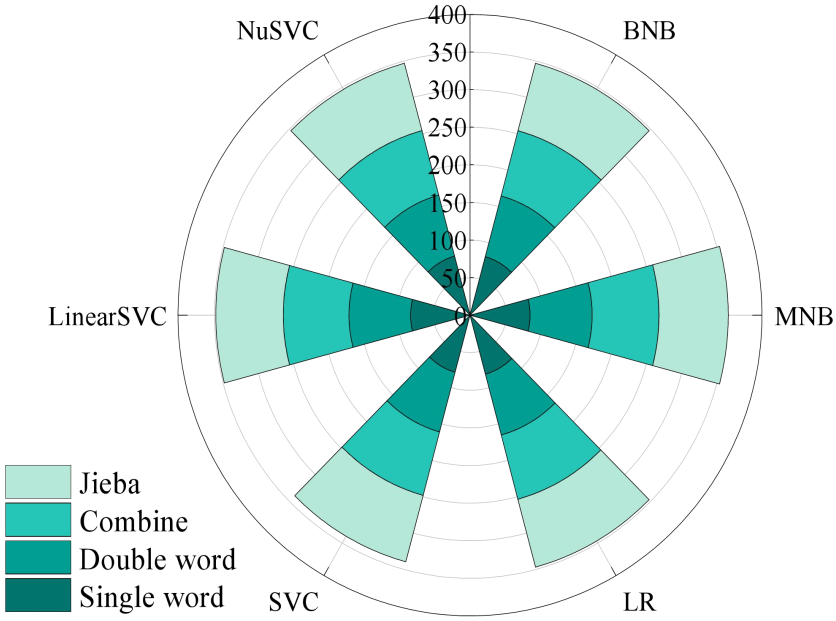

Figure 3. Accuracy of different classification methods.

In summary, three classifiers, MNB, LR and LinearSVC, are selected to detect the accuracy of the classifiers in different dimensions by using Jieba participles to select information-rich words as classification features, and the accuracy of the classifiers in different dimensions is shown in Figure 4. The overall classification accuracy of the classifier is higher when the feature dimension is 2500, while the classification accuracy of Multinomial Bayes (MNB) is higher when information-rich words are obtained as features using Jieba participles, which is 94.71%, and the accuracies of LR and LinearSVC at this time are 93.62% and 93.25%. When the feature dimension is 2500, the classifier is the most effective and the most appropriate features are selected, so the information-rich words obtained by using jieba participle combined with chi-square statistic, which are the most appropriate features, will be set as the candidate emotion words when the feature dimension is 2500.

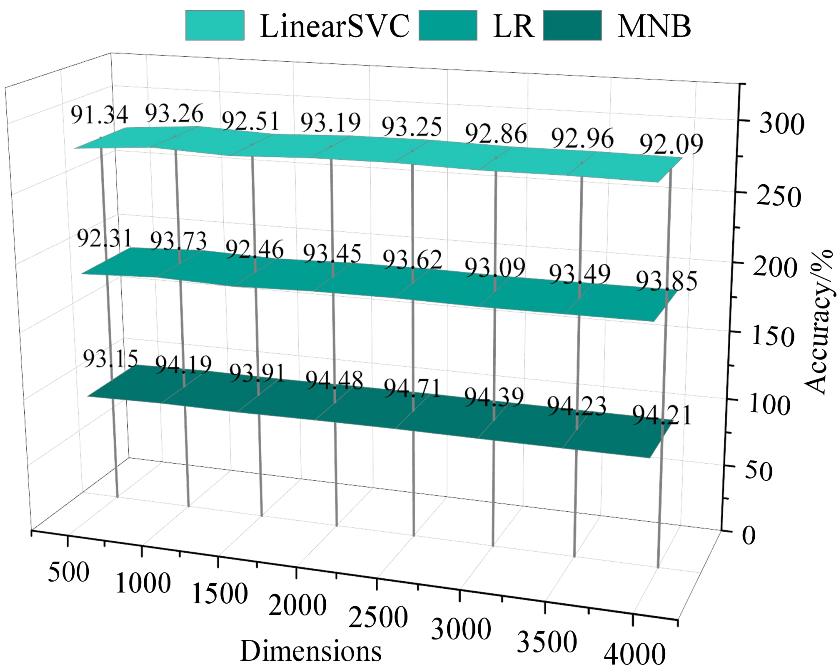

Figure 4. The accuracy of the classifier in different dimensions.

4.2.3. Algorithm comparison experiment

The text data in the dataset is selected for testing, and the emotion calculation model of this paper is applied to analyze it from the three dimensions of positive emotion, neutral emotion and negative emotion, and the results of the emotion tendency analysis are shown in Table 1. It can be seen that the emotion calculation model based on improved SO-PMI proposed in this paper has a high accuracy rate of 84.43%, 78.65% and 82.10% for positive emotion, neutral emotion and negative emotion, and it can be used to make a determination of the text emotional tendency.

Table 1. Emotional bias analysis results.

Affective tendency	Return result (number)	Correct result (number)	Accuracy/%
Positive emotion	1728	1459	84.43%
Neutral emotion	4286	3371	78.65%
Negative emotion	1257	1032	82.10%
Total	7271	5862	80.62%

In order to better analyze the improved SO-PMI method, the traditional SO-PMI method is used to compare with the improved SO-PMI algorithm, and at the same time, in order to determine the actual effect of the method in sentiment propensity analysis, the relatively mature SVM-based method is used to compare with it. The results of the comparison of emotional propensity classification are shown in Figure 5. Compared with the traditional SO-PMI method, the method in this paper has a large improvement, and the F1 values of positive sentiment, neutral sentiment, and negative sentiment are improved by 11.96%, 6.96%, and 6.89%, respectively. Compared with the SVM-based method to come, the method in this paper also has some advantages, and there is also a large improvement in the accuracy and recall, and the F1 value is improved by 2.88%, 4.30%, and 2.95%, respectively. It can be seen that the improved SO-PMI method proposed in this paper can effectively discriminate the emotional tendency of text, and is better than the traditional SO-PMI method and SVM-based method.

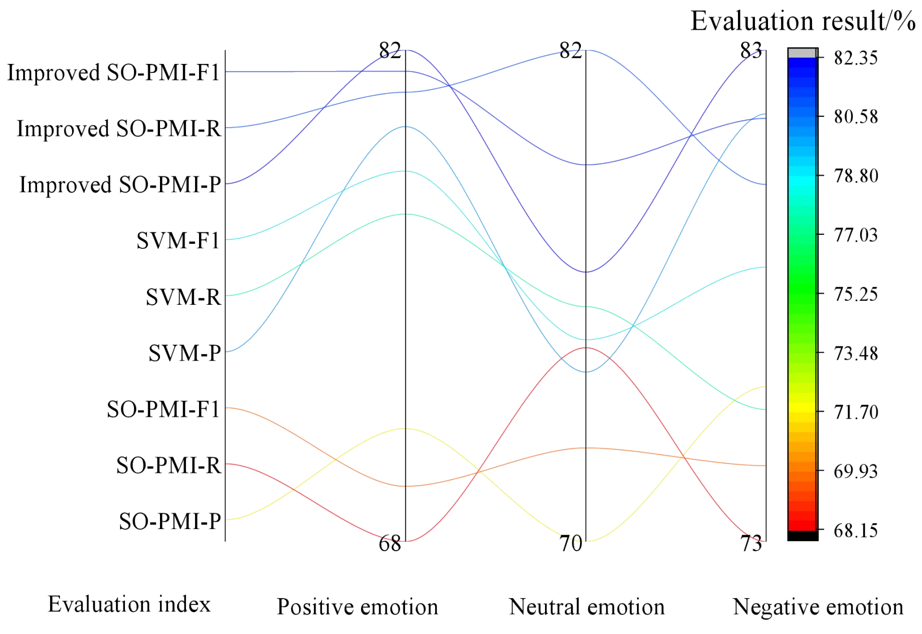

Figure 5. The contrast results of emotional bias classification.

4.2.4. Comparison of the degree of emotional disposition

In this experiment, the positive text and negative text will be divided into three grades each: average, strong, and very strong, with the value range of [0,5], [5,10], and [10, max], respectively, and the experiment will be carried out on 1,200 items of each of the positive text and negative text, which will be categorized into three grades according to the scores, and then the validity of the classification will be verified according to the experimental results. The positive emotional tendency classification results and negative emotional tendency classification results are shown in Figure 6 and Figure 7. The stronger the emotional tendency of the text, the higher its classification accuracy. The accuracy rate of all three emotional tendency classifications is above 86%, and it can be considered that the sentiment model based on the improved SO-PMI has a higher accuracy in emotional tendency segmentation and achieves the purpose of the experiment.

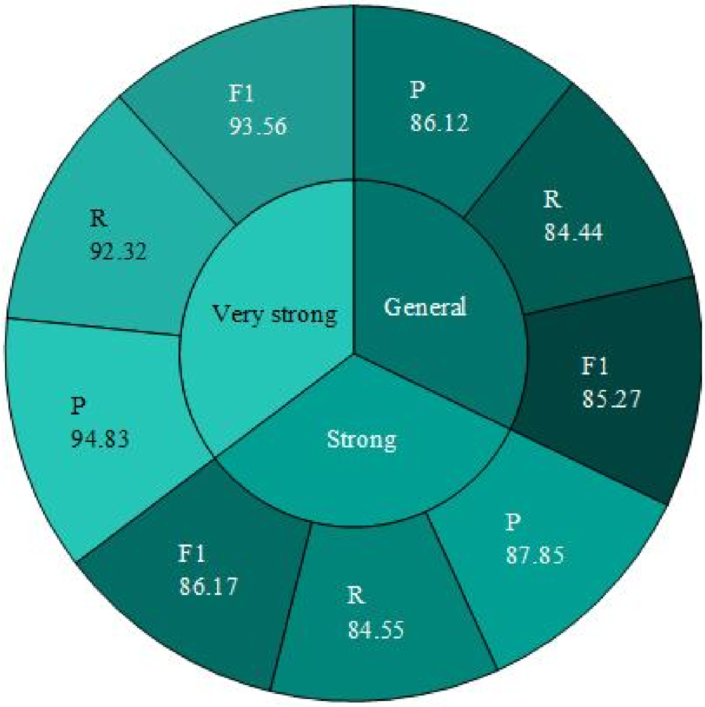

Figure 6. The classification results of positive emotional bias.

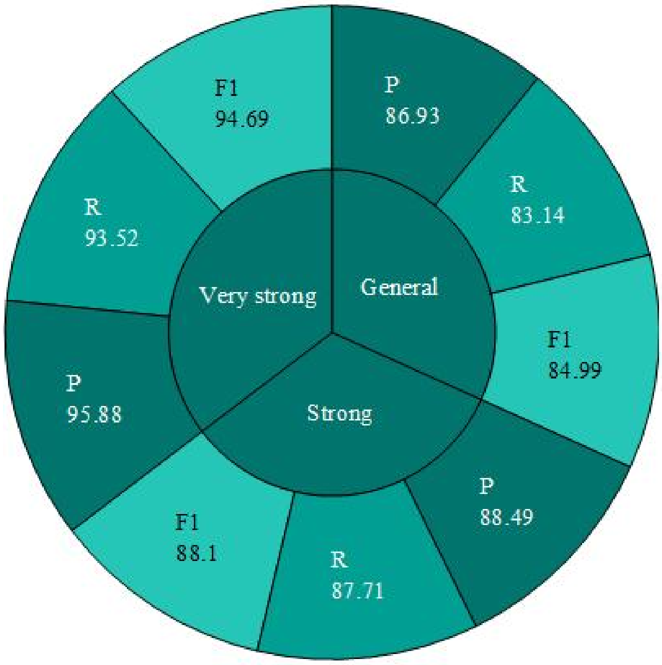

Figure 7. The classification results of negative emotional bias.

4.3. Analysis of empirical results

Based on the seven emotion types of the basic emotion lexicon, Emotion Vocabulary Ontology: sorrow, evil, good, shock, fear, joy, and anger, this paper's emotion computation model based on the improved SO-PMI was applied to emotionally analyze the collected textual data of the works of

Filipino-Chinese women writers, and the emotion classification results of the works' texts are shown in Figure 8. The percentages of sorrow, evil, good, shock, fear, joy and anger are 18.77%, 14.42%, 9.41%, 15.17%, 16.91%, 8.69% and 16.63%, respectively. The percentages of sorrow, fear and anger are the highest, while those of good and joy are the lowest, both below 10%, indicating that the female emotions conveyed in the works of Filipino Chinese female writers are mostly sorrow, fear and anger, which contain These include the authors' nostalgia for their homeland, sorrow for the dispersal of their families, criticism of the aesthetic value of traditional concepts, and sympathy for and reflection on the female protagonists in their works.

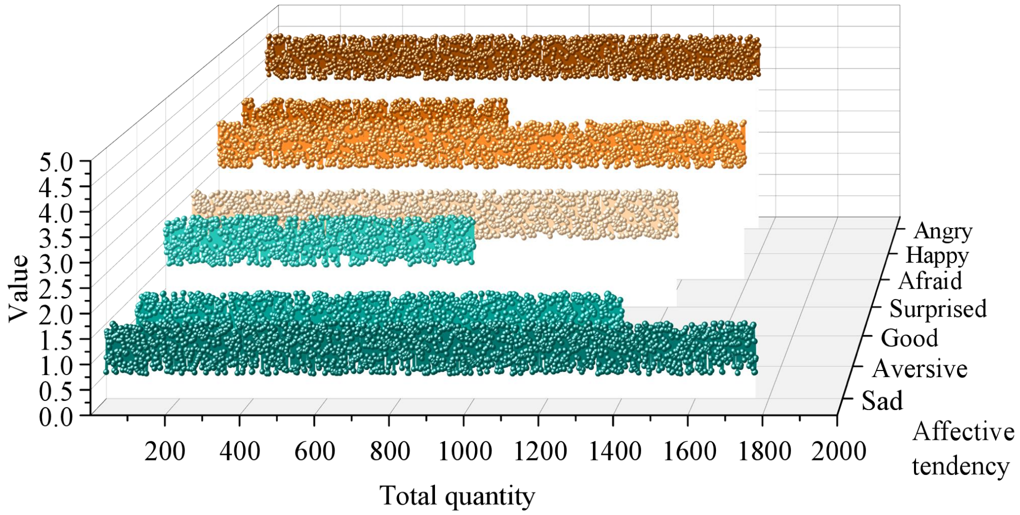

Figure 8. The emotional classification of the work text.

5. Conclusion

Literary Sentiment Calculation provides a tool for quantitative analysis, enabling researchers to explore the emotional patterns embedded in literary works in an unprecedented way. In this paper, we implement sentiment computation of texts through sentiment lexicon construction using the improved SO-PMI algorithm, and after experimentally analyzing the performance of the model, we conduct a sentiment analysis of the works of Filipino Chinese female writers. The main research results are as follows:

(1) Compared with a single sentiment dictionary, the sentiment dictionary constructed in this paper has higher evaluation index results, and its sentiment classification F1 values in positive and negative texts are 72.63% and 70.43%, which proves the validity and practicality of the constructed sentiment dictionary. Meanwhile, the improved SO-PMI algorithm has a large improvement over the traditional SO-PMI algorithm, and the F1 values of positive sentiment, neutral sentiment and negative sentiment have been improved by 11.96%, 6.96% and 6.89%, respectively. The effectiveness of the modified SO-PMI algorithm for sentiment lexicon expansion is illustrated.

(2) In the emotion analysis of the works of Filipino Chinese female writers, the emotions of sadness, fear, and anger account for the highest proportion, 18.77%, 16.91%, and 16.63%, respectively, while the emotions of goodness and joy account for a smaller proportion, which indicates that the expression of emotions in the works of Filipino Chinese female writers is mostly sadness, sorrow, and anger, which is in line with the background of cross-cultural impact and the environment of the works' creation in which the development of women's consciousness is taking place.

It has been an inevitable trend to conduct research on overseas Chinese literature from the perspective of digital humanities. This paper adopts an affective measurement model to study the emotional expression of Filipino Chinese women writers' works, and its study of Chinese literature provides a new cultural research perspective, and the two-way penetration of digital research and cultural research will provide new possibilities for their respective studies.

References

1. Jocson, J. V. (2020). A feminist reading of Filipino women poets. Rupkatha Journal on Interdisciplinary Studies in Humanities, 12(6), 1-12.
2. Heruela, M. (2016). The Filipino Chinese Woman: Creation of Transnational Feminist Counterpublic in Modern Philippine Chinese Women's Short Fiction. Proceedings of Asian Studies 2016, 99.
3. Dela Cruz, N. L. (2015). Who counts as Filipino? Philosophical issues of identity and the chinese filipino. Budhi: A Journal of Ideas and Culture, 19(2), 3.
4. Stenberg, J. (2023). Diverse fragility, fragile diversity: Sinophone writing in the Philippines and Indonesia. Asian Ethnicity, 24(1), 59-77.
5. Camba, A. J., & Lung, S. (2021). Chinese capital as a cultural object: self-identification and Filipino-Chinese discourses on sinicization, brokerage, and distinction. Translocal Chinese: East Asian Perspectives, 15(2), 186-213.
6. Gonzales, W. D. W. (2017). Language contact in the Philippines: The history and ecology from a Chinese Filipino perspective. Language Ecology, 1(2), 185-212.
7. Cheung, K. K. (2015). Chinese and Chinese American Life-Writing. Cambridge Journal of China Studies 10.2 (2015): 1-20, 10(2), 1-20.
8. Cruz, C. (2017). The (mis) education of the Filipino writer: The Tiempo Age and institutionalized creative writing in the Philippines. Kritika Kultura, 1(28), 3.
9. Mendoza, M. L. (2012). Filipino Women Writers in English and the Work of Apprenticeship. WorkingUSA, 15(1), 103-119.
10. Vardhan, P. (2018). Feminism and women writers in English. IMPACT: International Journal of Research in Humanities, Arts and Literature (IMPACT: IJRHAL), 6.
11. Koussouhon, A. L., Akogbeto, P. A., & Allagbe, A. A. (2015). Portrayal of male characters by a contemporary female writer: A feminist linguistic perspective. International Journal of Advanced Research, 3(12), 314-322.
12. Spacks, P. M. (2022). The female imagination: A literary and psychological investigation of women's writing. Routledge.
13. Howell, S. (2015). The evolution of female writers: An exploration of their issues and concerns from the 19th century to today. University of Hawaii at Hilo HOHONU, 13.
14. Ingalls, V. (2020). Who creates warrior women? An investigation of the warrior characteristics of fictional female heroes based on the sex of the author. Evolutionary behavioral sciences, 14(1), 79.
15. Qian, Y. (2019, July). Gender stereotypes differ between male and female writings. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop (pp. 48-53).
16. Bijami, M., Kashef, S. H., & Khaksari, M. (2013). Gender Differences and Writing Performance: A Brief Review. International Journal of Education and Literacy Studies, 1(2), 8-11.
17. Zhang, M., Bennett, R. E., Deane, P., & van Rijn, P. W. (2019). Are there gender differences in how students write their essays? An analysis of writing processes. Educational Measurement: Issues and Practice, 38(2), 14-26.
18. Anjarwati, R., Setiawan, S., & Laksono, K. (2021). Experiential meaning as meaning making choice in article writing: A case study of female and male writers. Heliyon, 7(4)
19. Shamina, N. V. (2022). THE SPECIFICS OF THE FEMALE THEME INTERPRETATION IN VICTORIAN WOMEN WRITERS'WORKS. Russian Linguistic Bulletin, (8 (36)), 3.
20. Chakraborty, P., & Dasgupta, S. (2018). Role of creative writing on creative personality, quality of life, meaning in life and affect balance of male and female creative writers. Indian Journal of Community Psychology, 14(2).
21. Gubar, S. (2014). " The Blank Page" and the Issues of Female Creativity. Lectora (Barcelona), (20), 0249-269.
22. Annas, P., & Peseroff, J. (2015). A feminist approach to creative writing pedagogy. Creative writing pedagogies for the twenty-first century, 78-101.