

Research Article

Analysis of Poetry Style Based on Text Classification Algorithm

Can Wang 

Basic Theory Department, Henan Industry and Trade Vocational College, Zhengzhou 450000, China

Correspondence should be addressed to Can Wang; wangcan@hngm.edu.cn

Received 8 December 2021; Revised 26 December 2021; Accepted 5 January 2022; Published 31 January 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Can Wang. 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.

In order to realize the effect of intelligent poetry style analysis, this paper applies the text classification algorithm to the poetry style analysis, combines the knowledge representation algorithm to perform text classification and recognition, improves the algorithm, and applies it to the poetry style analysis model. Moreover, this paper combines intelligent algorithms to construct a poetry style analysis system, constructs the system's functional modules, preprocesses the poetry documents in the corpus, and maps them to the vector space that can be directly processed by the computer. In addition, after constructing the system model, this paper verifies the poetry style analysis system based on the text classification algorithm through simulation experiments. From the research results, the effect of the poetry style analysis method based on the text classification algorithm proposed in this paper is very good, which meets the actual needs of poetry style analysis.

1. Introduction

At present, a lot of progress has been made in natural language processing technology. As a branch of natural language processing, computer processing of literary language is now in front of academia. With the support of the National Natural Science Foundation of China, this paper is trying to make preliminary explorations in this field. The focus of this paper is how to use machines to realize the style of poetry. As one of the essences of Chinese traditional culture, the classical poems have meaningful meaning, simple and deep sentiment, and their style is intoxicating. However, for a long time, the judgment of the style of poetry works is mostly based on the reader's judgment and cognition of the style based on experience and feeling. Moreover, traditionally, there are no clear quantitative rules and no formal rules. For machines, it can only rely on the text content of the poem to determine the style. From the form, the poem is also a paragraph of text, and the judgment of its style can also be regarded as the classification of the text style. Therefore, this problem is essentially transformed into a text classification problem in machine learning. That is, under a given style classification system, the machine automatically learns the rules according to the content of the text, and establishes a classification discriminator according

to the rules, and the classification discriminator automatically determines the style of the text. Based on this, this paper proposes a model framework for computing the style of Chinese classical poetry based on machine learning. This research is of great significance for expanding the application field of machine learning, for the informatization and excavation of traditional poetry, and for the language understanding of poetry and the computer-aided research of various literary works.

As one of the treasures in the history of Chinese literature, poetry has been passed down through the ages because of its enduring charm. This literary form, favored by the general public and literati and poetas, has been growing, evolving, and disseminating and has reached the peak of its development in the Tang and Song dynasties. The research on Tang poetry has always been a research hotspot among scholars and literary lovers. Since ancient times, countless experts and scholars have carried on extensive and in-depth research on Tang poetry and poetry from a literary perspective. However, with the continuous development of information science and technology, people have begun to try to use information science and technology to deal with some of the language problems encountered in work and life. Natural linguistics has developed rapidly in recent years. Faced with the victories of natural language processing in

modern Chinese, some natural language processing methods are applied to ancient poetry. At the same time, in view of the difference between ancient poetry and modern Chinese, from the perspective of computer, integrated data mining, genetic algorithm, and other technologies and resources such as HowNet, it focuses on the application of text classification technology in the classification of poetry style.

In the process of creating poems, the different talents, temperaments, literary accomplishments, life experience, ideological realm, and ideological feelings of the literati and the ink guest will affect the content and expression of the poems written by the literati and ink guest, and they will also have the ability to Poems with different characteristics, therefore, do not have the same style of poems. However, in the process of appreciating classical poems, if you blindly pay attention to the individual differences between poems and ignore the common characteristics of the poems, it is likely to cause the problem of nihilism.

This article combines the intelligent computer algorithm to apply the text classification algorithm to the poetry style analysis and builds an intelligent classification system, which lays the foundation for the intelligent classification of poetry.

2. Related Work

Using computer technology to assist in the study of the field of classical poetry literature, some scholars have begun preliminary attempts and explorations early on. The construction of the classical poetry database and cultural knowledge base is in the computer research laboratories of some universities and research organizations [1]. With the increasing progress of natural language processing technology in the study of Chinese language, it can even gradually support the analysis of rhythm, emotional element, and style characteristics of classical poetry [2]. “Computational Poetics” is proposed in this context [3]. There have been many research results on the related fields of poetry style. According to the literature [4] “The topic model-based corpus construction and computer-aided creation research,” the topic model based on LDA uses the method of reference word recommendation to study and analyze the word characteristics in poetry, vocabulary semantic analysis, and style feature analysis. Literature [5] uses the vector space model (VSM) to represent the text of poems and proposes two types of classification models of classical poems, bold and graceful and graceful. The classification models are based on machine learning and natural language processing. Poetry styles are classified, and a good judgment effect has been achieved. According to literature [6] after segmenting the poems according to the poetry metric and statistical word segmentation methods, they established a poetry vocabulary database and used machine learning methods to analyze the characteristics of poetry styles, classification evaluation, and sentiment analysis. In addition, research in related fields has focused on the algorithm improvement of the poetry style machine classification model, in order to gradually improve the accuracy of the classification model for the poetry style evaluation. The early stage corpus construction and computer processing

procedures have been very clear. Various research work is still in-depth [7]. It can be seen that there have been quite a lot of research studies on the classification and judgment of poetry style, which has accumulated rich experience for the classification and judgment of poetry style based on machine learning in this article.

The handling of poems can actually be regarded as a text handling problem with a special content in form [8]. Literature [9] developed a “Poetry Computer Aided Research System,” which takes the vocabulary in the poem as the main research unit to realize word retrieval, word frequency statistics, and image indexing. Literature [10] proposed a natural language processing technology based on word connection and used it for the understanding of poetry language. It successfully carried out a test of poetry vocabulary material labeling, a preliminary analysis test of poetry language, and an evaluation test of poem language boldness and graceful style. Literature [11] combined naive Bayes with the genetic algorithm, and put forward a calculation model for judging the bold and graceful style of classical poetry based on each character of the poem as the smallest unit, and tested the corpus in the classic poetry. Good results have been achieved. Literature [12] introduced Bayesian classification technology into poetry research from the subject point of view and achieved good experimental results. Literature [13] proposed a classification method of poem topics based on point mutual information and LDA and achieved a good classification effect on the relationship between poem topics and the evolution of the times. Literature [14] uses the vector space model (VSM) to convert the poetry text into a vector and selects the word feature through the chi-square test. Finally, the text classifier is constructed based on the naive Bayes and support vector machine algorithm, and the average classification accuracy rate reaches 74.7%. With the rise of deep learning, the literature [15] applied the RNN neural network to the generation of ancient Chinese poetry, and used the entire historical corpus of poetry as training corpus, adding certain constraints between the words and lines of the generated poetry to achieve it has a better effect than the traditional poetry generation system. Literature [16] uses the RNN model to generate poetry quatrains, and successfully adds an attention mechanism, so that the poetry can learn semantics, structure, rhyme, and other information at the same time. Literature [17] based on the RNN model, by optimizing the word vector of poetry, introducing attention mechanism and hybrid training, a model that can produce topic-related poems based on keywords has been obtained, and good results have been achieved.

3. Knowledge Representation of Intelligent Poetry Text Generation System

Like many other intelligent systems, knowledge representation is a very important key decision during system construction. We treat Chinese characters and works of art as images in a parametric form. Mathematically, the “image” of Calligraphy C is represented as a band-shaped image area covered by a series of ellipses. It means that this series of C is

composed of num_0 ellipses $P_1, P_2, \dots, P_{\text{num}_0}$. $F_0 = \{1, 2, \dots, \text{num}_0\}$, (x_i, y_i) , and a_i, b_i are used as the center, semimajor axis, and segment semiaxis of ellipse P_i , respectively. We define the text of traditional Chinese poetry as

$$\text{Area} \triangleq \left\{ (x, y) \in R^2 \mid \exists i \in F_0, \frac{(x - x_i)^2}{a_i^2} + \frac{(y - y_i)^2}{b_i^2} \leq 1 \right\}. \quad (1)$$

The above knowledge representation method is inspired by the method of the paper. In this paper, a band-shaped area is defined as an ellipse moving along a predefined curve.

We now turn to the concept of equivalence and use a formalized definition of the hierarchical representation of the text of poetry. If R is an equivalence relation defined in the $A = \{a_1, a_2, \dots, a_n\}$ field, that is to say, R is as follows [18]:

- (1) Reflexive,
- (2) symmetrical, and
- (3) passed.

Domain A can be decomposed into a series of subsets A_1, A_2, \dots, A_m in relation R , which satisfy the following properties:

- (a) If $i \neq j, 1 \leq i, j \leq m$, then $A_i \cap A_j = \phi$ and
- (b) $\forall a_i \in A, \exists j, 1 \leq j \leq m, \text{ s.t. } a_i \in A_j$.

Under the equivalence relation R , if $(a_i, a_j) \in R, 1 \leq i, j \leq n$, we say that a_i is equivalent to a_j , which can be expressed as $a_i \leftrightarrow_R a_j$.

In order to establish the hierarchical representation of poetry text, we introduced four equivalence relations R_1, R_2, R_3, R_4 :

R_1 : all the structural ellipses used to form the same original sign are equivalent to each other.

R_2 : all the original strokes used to form the same combination sign are equivalent to each other.

R_3 : all the combined strokes used to form the same radical are equivalent to each other.

R_4 : all the radicals used to form the same Chinese character are equivalent to each other.

We assume that the constructed ellipse $C = \{Pt \mid t \in F_0\}$ is decomposed into num_1 equivalence classes under the equivalence relation R_1 , which can be expressed as $F_1 \triangleq \{1, 2, \dots, \text{num}_1\}$. The num_1 original signs are decomposed into num_2 equivalence classes under the equivalence relation R_2 , which can be expressed as $F_2 \triangleq \{1, 2, \dots, \text{num}_2\}$. num_2 combined strokes are decomposed into num_3 equivalence classes under equivalence relation R_3 , which can be expressed as $F_3 \triangleq \{1, 2, \dots, \text{num}_3\}$. The num_3 radicals are decomposed into num_4 equivalence classes under the equivalence relation R_4 , which can be expressed as $F_4 \triangleq \{1, 2, \dots, \text{num}_4\}$ [19].

In other words, in a specific poetic text work C , there are a total of num_1 original strokes $P_{1,i}, i \in F_1$. Moreover, C contains num_2 combined strokes $P_{2,i}, i \in F_2$. We can say that C has num_3 radicals $P_{3,i}, i \in F_3$. We can also say that C consists of num_4 words $P_{4,i}, i \in F_4$. Obviously, when we use the granularity of a single word to apply comprehensive reasoning, $\text{num} \equiv 1$ [20].

Now, the hierarchical representation of traditional poetry text works can be formally described as follows:

$$\left\{ \begin{array}{l} \mathbf{P}_{k,1} = \left\{ \mathbf{P}_{k-1,s} \mid \mathbf{P}_{k-1,s} \xleftrightarrow{R_k} \mathbf{P}_{k-1,1}, \quad s \in F_{k-1} \right\} \\ \mathbf{P}_{k,l} = \left\{ \mathbf{P}_{k-1,s} \mid \mathbf{P}_{k-1,s} \xleftrightarrow{R_k} \mathbf{P}_{k-1,q} \text{ in which } q = \min \left(t \mid \mathbf{P}_{k-1,t} \notin \bigcup_{i=1}^{l-1} \mathbf{P}_{k,i}, t \in F_{k-1} \right), s \in F_{k-1} \right\}, \quad l \in \frac{F_k}{\{1\}} \quad k = 1, 2, 3, 4 \\ \mathbf{P}_{0,l} = P_l, \quad l \in F_0 \end{array} \right\}, \quad (2)$$

where the number of elements in the set M is $|M|$. Therefore, the following hierarchical relationship expressed in poetry text works can be expressed as follows:

$$\left\{ \begin{array}{l} \sum_{s=1}^{|F_k|} |\mathbf{P}_{k,s}| = |F_{k-1}| = \text{num}_{k-1}, \quad k = 1, 2, 3, 4, \\ |\mathbf{P}_{0,s}| = 1, \quad s = 1, \dots, |F_0|. \end{array} \right\} \quad (3)$$

This hierarchical representation describes how the poetry text works at the lowest level are composed of structural ellipses. Each higher level describes how to generate a representation of its current level from the information of

the lower level. In essence, it is a tree-shaped knowledge representation.

At the 0th level of the hierarchical representation, the poetry text works are regarded as a series of ellipses, denoted as P_0 . These ellipses are called structural ellipses for poetry text works. For each construction ellipse, it is denoted as $P_{0,i}$, and (x_i, y_i) is taken as its center, and a_i and b_i are taken as its semimajor axis and semiminor axis, respectively. Then, the image of the poetic text work C will be presented as a series of image areas covered by an ellipse.

If a lower-level element cannot be merged with other elements at the same level, it will be promoted to an adjacent

higher-level element. In the same way, it is also possible that a radical will be downgraded into a combined gesture, or even an original one. Through these upgrade and downgrade arrangements, we have obtained a unified 6-level hierarchical system, which will be used in the generation of poetry text.

We denote the first structural ellipse of the poem text work as $P_{k,s}$, and its corresponding topological structure is $T_{k,s}$. $k = 1, 2, 3, 4, s \in F_k \circ T_{k,s}$ contains topology information,

$$T_{k,s} \triangleq (\text{TCR}_{k,s}, \text{TCS}_{k,s}) \triangleq \left(\begin{bmatrix} TR_{k,1+l} \\ TR_{k,2+l} \\ \vdots \\ TR_{k,|P_{k,s}|+l} \end{bmatrix}, ([TS_{k,1+l} \quad TS_{k,2+l} \quad \cdots \quad TS_{k,|P_{k,s}|+l}] \right). \quad (4)$$

Here, $l = \sum_{i=1}^{s-1} |P_{k,i}|$; $s \in F_k$; $k = 1, 2, 3, 4$. $\text{TCR}_{k,s}$ and $\text{TCS}_{k,s}$ is the scaling and transition matrix, respectively. The strict mathematical definition of matrix $TR_{k,z}$, $TS_{k,z}$ is as

$$\left\{ \begin{array}{l} TR_{k,z} \triangleq \begin{bmatrix} RW_{k,z} & 0 \\ 0 & RH_{k,z} \end{bmatrix} \triangleq \begin{cases} \begin{bmatrix} \frac{CW_{k,s}}{CW_{k-1,z}} & 0 \\ 0 & \frac{CH_{k,s}}{CH_{k-1,z}} \end{bmatrix}, & (k = 2, 3, 4), \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, & (k = 1), \end{cases} \\ TS_{k,z} \triangleq [RX_{k,z} \quad RY_{k,z}]^T \triangleq \begin{cases} \left[\frac{CX_{k-1,z} - CX_{k,s}}{CW_{k,s}} \quad \frac{CY_{k-1,z} - CY_{k,s}}{CH_{k,s}} \right]^T, & (k = 2, 3, 4), \\ [0 \ 0]^T, & (k = 1). \end{cases} \end{array} \right. \quad (5)$$

If $k \geq 1, P_{k,i}$ must be a combination of one or more construction elements one level lower. We call the latter substructural elements. All the information needed to combine $P_{k,i}$ is stored in $T_{k,i}$, which is the topology constructor of $P_{k,i}$.

Through the topology constructor, we can build a one-to-one mapping of different levels of hierarchical representation. In other words, any hierarchical representation pixel $[x_{k,s} \ y_{k,s}]^T$ of the s -th structural element $P_{k,s}$

and through it, $P_{k-1,1+l}, P_{k-1,2+l}, \dots, P_{k-1,|P_{k,s}|+l}$ ($l = \sum_{i=1}^{s-1} |P_{k,i}|$) can be used to form $P_{k,s}$.

We denote the bounding box of the image area occupied by element $P_{k,s}$ as $C_k \cdot s = [CX_{k,s} \ CY_{k,s} \ CH_{k,s} \ CW_{k,s}]$, $k = 1, 2, 3, 4, s \in F_k$. Here, $CX_{k,s}, CY_{k,s}, CH_{k,s}, CW_{k,s}$ are the x, y coordinates of the lower left corner of the bounding box $C_{k,s}$ and its height and width, respectively. All coordinates are in the world coordinate system. Therefore, we have

belonging to the k -th level can be uniquely mapped to the hierarchical representation pixel $[x_{1,t} \ y_{1,t}]^T$ of the c -th structural element $P_{1,t}$ of the first level. Among them, $l > k, 1, k = 0, 1, 2, 3, 4, \exists m_j \in F_{k+j}$ ($j = 1, 2, \dots, 1 - k - 1$) s.t. $P_{k,s} \in P_{k+1,m_1} \in \dots \in P_{k+(1-k-1),m_{1-k-1}} \in P_{1,t}$.

If we introduce the operator $\nabla_{m,n}^b$,

$$\nabla_{m,n}^b \begin{bmatrix} x \\ y \end{bmatrix}_{2 \times 1} \triangleq TR_{m+1,n} \begin{bmatrix} x \\ y \end{bmatrix}_{2 \times 1} + TS_{m+1,n}. \quad (6)$$

Then, there will be a relationship:

$$[x_{l,t} \ y_{l,t}]^T = \nabla_{l-1, m_{l-k-1}}^b (\dots (\nabla_{k+1, m_1}^b (\nabla_{k,s}^b [x_{k,s} \ y_{k,s}]^T))) \quad (7)$$

The s -th structural element of the k -th level of the hierarchical system is denoted as $P_{k,r,s}$, and its parameter is

denoted as $E_{k,s}$, $k = 0, 1, \dots, 4$, $s \in F_k$. We now introduce a new matrix operator ∇_n^c , which can be used to generate an $m \times \sum_{l=1}^n d_l$ -dimensional matrix $M = (a_{i,j})_{m \times \sum_{l=1}^n d_l}$ and n $m \times d_l$ ($l = 1, 2, \dots, n$)-dimensional matrices $M_1 = (a_{l,i,j})_{m \times d_1}$ ($l = 1, 2, \dots, n$). If $\nabla_n^c(M_1, M_2, \dots, M_n) = M$, then we have

$$a_{i,j} = \begin{cases} a_{z+1,i,j} - \sum_{t=1}^z d_t \text{ in which } \sum_{t=1}^z d_t < j \leq \sum_{t=1}^{z+1} d_t, & z = 1, 2, \dots, n-1, \\ a_{1,i,j}, & j \leq d_1, \end{cases} \quad (8)$$

$$i = 1, 2, \dots, m.$$

We further define a matrix operator ∇_n^d :

$$\nabla_n^d(A) \triangleq \nabla_n^c \underbrace{(A, A, \dots, A)}_n. \quad (9)$$

Then, the hierarchy of poetry text works can be formally expressed as follows:

$$\begin{cases} E_{0,i} = \begin{bmatrix} x'_i & y'_i \\ E_{k,s} \end{bmatrix} = \nabla_{|P_{k,s}^c|} \left(\nabla_{k-1,1+l}^e E_{k-1,1+l}, \nabla_{k-1,2+l}^e E_{k-1,2+l}, \dots, \nabla_{k-1,|P_{k,s}^c|+l}^e E_{k-1,|P_{k,s}^c|+l} \right). \end{cases} \quad (10)$$

Among them,

$$l = \sum_{i=1}^{s-1} |P_{k,i}|, \quad k = 1, 2, 3, 4, s \in F_k. \quad (11)$$

At this time, the operator is defined as follows:

$$\nabla_{n,m}^e E_{n,m} \triangleq \nabla_2^c \left(\left(\nabla_2^c (TR_{n,m}, 0_{2 \times 2}) \right)^T, \left(\nabla_2^c (0_{2 \times 2}, TR_{n,m}) \right)^T \right) E_{n,m} + \nabla_{\text{col}(E_{n,m})}^d \left(\left(\nabla_2^c \left((TS_{n,m})^T, [0 \ 0] \right) \right)^T \right). \quad (12)$$

Among them, $\text{col}(E_{n,m})$ is the number of columns of matrix $E_{n,m}$, and $0_{2 \times 2}$ is a 2×2 -dimensional zero matrix.

In the above equation, each $E_{0,i}$ represents an area covered by the constructed ellipse, and

$((x - x'_i)^2/a_i'^2) + ((y - y'_i)^2/b_i'^2) \leq 1$. This ellipse is the normalized version of the original ellipse $((x - x_i)^2/a_i^2) + ((y - y_i)^2/b_i^2) \leq 1$ for its bounding box. That is, if $P_{0,i} \in P_1$ and $j, i \in F_0$, $j \in F_1$ is assumed, we have

$$\begin{bmatrix} x'_i \\ y'_i \\ a'_i \\ b'_i \end{bmatrix} = \begin{bmatrix} \frac{x_i - CX_{1,j}}{CW_{1,j}} \\ \frac{y_i - CY_{1,j}}{CH_{1,j}} \\ \frac{a_i}{CW_{1,j}} \\ \frac{b_i}{CH_{1,j}} \end{bmatrix}. \quad (13)$$

For an element higher than the 0th level, such as $P_{k+1,1}$, for all $k > 0$, it is composed of $N_{k+1,1}$ adjacent lower-level elements: $P_{k,11}, P_{k,12}, \dots, P_{k,1N_{k+1,1}}$. The corresponding shape matrix $E_{k+1,1}$ is deduced, which is obtained from the matrix $E_{k,11}, E_{k,12}, \dots, E_{k,1N_{k+1,1}}$ by linking columns by column. Because the parameterized representation of constructing an ellipse is a 4×1 -dimensional matrix, the above-mentioned recommended connection method will accurately generate a matrix with only 4 rows at a higher level. Each row of the generated matrix is called the parameterized representation domain of the element. The different domains of an element can be inferred independently.

The weight associated with the s -th field of the i -th source knowledge on the k -th level; that is, the weight of the i -th training example ($p_{k,l}^i$) is denoted as $w_{i,s}^i$. $w_{i,s}^i$ ($s = 1, \dots, 4$) is the reasoning strength of $p_{k,l}^i$ in synthetic reasoning. Therefore, the set of similar reasoning strengths arranged in order constitutes the "view sequence" of the current comprehensive reasoning process. This is what we said that different sequences of showing training examples will lead to different results of poetry text.

For the comprehensive reasoning process generated by novel poetry text works, we assume that n comprehensive source knowledge $P_{k,11}, P_{k,12}, \dots, P_{k,1n}$ at the k -th level participates in this reasoning process. Moreover, each comprehensive source knowledge consists of m components, where $\forall i \neq j; i, j = 1, 2, \dots, n \Rightarrow l_i, l_j \in F_k, l_i \neq l_j$. Therefore, we can use an $m \times 1$ -dimensioned matrix to represent a single comprehensive reasoning source and use an $m \times \sum_{i=1}^n l_i$ -dimensioned matrix to represent all active sources in the current comprehensive reasoning. If we apply synthetic reasoning at the single-character level, then $m \equiv 4$. The t -th attribute of source knowledge is denoted as $P_{k,li,t}$, which has an associated strength $W_{k,li,t}$ in the process of comprehensive reasoning, and they have the following relationship:

$$\sum_{i=1}^n \omega_{k,li,t} = 1 (t = 1, 2, \dots, m). \quad (14)$$

In the intelligent poetry text generation system, this reasoning intensity can be adjusted through a graphical user interface. The new knowledge generated in the synthetic reasoning process is denoted as $P_{k,r}$, and its matrix form parameterization is denoted as $E_{k,r}$. To apply

the principle of comprehensive reasoning process, all source knowledge in reasoning must have equal dimensions. That is to say, if we apply synthetic reasoning on source knowledge P_k, s and $P_{k,t}$, their corresponding matrix form parameterized representations $(E_{k,s})_{m_1 \times n_1}$ and $(E_{k,t})_{m_2 \times n_2}$ must satisfy the relationship $n_1 = n_2$. This equal requirement is a soft constraint of the synthetic reasoning process. If it is violated, we can introduce a source knowledge equality operator ∇_t^e to relax this soft constraint. This is very similar to the mapping problem in similar reasoning. Next, we will introduce a few symbols and define ∇_t^e in a strict mathematical way.

We first denote $en, i \triangleq [\sigma(i, 1) \sigma(i, 2) \dots \sigma(i, n)]^T$, where $\sigma(i, j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$.

We can also obtain a discrete curve composed of the centers of all the covering ellipses for each structural element in the intelligent poetry text generation system, which can be expressed as P_k . Its corresponding parameterization is expressed as E_k, l_s . Therefore, a dispersion curve acquisition method can be described as follows:

$$\begin{aligned} C_{k,l_s} &= \left(\nabla_2^c \left((e_{4,1}^T \times E_{k,l_s})^T, (e_{4,2}^T \times E_{k,l_s})^T \right) \right)^T \\ &= \begin{bmatrix} x_1 & x_2 & \dots & x_{co} \\ y_1 & y_2 & \dots & y_{co} \end{bmatrix}, \quad co = \text{col}(E_{k,l_s}). \end{aligned} \quad (15)$$

Then, an overview of the key columns of a matrix can be introduced and defined. If a curve C_k has $v + 1$ key points and its coordinates are $C_{k,1se_{co,u0}}, C_{k,1se_{co,u1}}, \dots, C_{k,1se_{co,u v}}$, respectively, then the key column of E_{k,l_s} can be identified as $E_{k,1se_{co,u0}}, E_{k,1se_{co,u1}}, \dots, E_{k,1se_{co,u v}}$. Aiming at the key point extraction of a given plane curve C_{k,l_s} , we use the algorithm introduced in this paper. Moreover, more refined key point extraction algorithms are introduced in, but they may cause more serious computational overhead.

We assume that now we have identified $v + 1$ key columns, which are the u_0, u_1, \dots, u_v -th matrix column, through the parameterized representation E_{k,l_s} of a matrix form of a comprehensive reasoning source P_{k,l_s} , where $1 = u_0 < u_1 < \dots < u_v = n = \text{col}(E_{k,l_s}), s = 1, 2, \dots, n$. Therefore, we can define the capacity equality operator of the comprehensive reasoning source ∇_t^e as follows:

$$\left(\nabla_t^e (E_{k,l_s}) \right) e_{t,i} \triangleq E_{k,l_s} e_{\text{col}(E_{k,l_s}), \theta^i}, \quad i = 1, 2, \dots, t. \quad (16)$$

Among them,

$$\begin{aligned} \theta &= \left[u_j + \frac{u_{j+1} - u_j}{[t \times (j+1)/v] - [t \times j/v]} \left(i - \left[\frac{t \times j}{v} \right] \right) \right]; \\ \left[\frac{t \times j}{v} \right] &< i \leq \left[\frac{t \times (j+1)}{v} \right]. \end{aligned} \quad (17)$$

$j \in \{0, 1, \dots, v-1\}; s = 1, 2, \dots, n$. $[\]$ is an integer trimming operator. Specifically, if each column of matrix E_{k,l_s} is considered to be a key column, we can use a simpler mapping to define the operator ∇_t^e :

$$\left(\nabla_t^e(E_{k,l_s})\right)e_{t,i} \triangleq E_{k,l_s} e_{\text{col}(E_{k,l_s}), [\text{ixcol}(E_{k,l_s})/t]}. \quad (18)$$

Among them, $i = 1, 2, \dots, t$, and $[\]$ is an integer trim function. Based on the definition of ∇_t^e , we can further define a comprehensive inference source knowledge superposition operator ∇_t^f :

$$\begin{aligned} \nabla_n^f(M_1, M_2, \dots, M_n) &\triangleq \nabla_n^c(\nabla_h^e(M_1), \nabla_h^e(M_2), \dots, \nabla_h^e(M_n)), \\ h &= \max\{\text{col}(M_i) | i = 1, 2, \dots, n\}. \end{aligned} \quad (19)$$

Through the operator ∇_t^f , we can obtain the matrix representation of the superimposed comprehensive reasoning source:

$$E_k \triangleq \nabla_n^f(E_{k,l_1}, E_{k,l_2}, \dots, E_{k,l_n}). \quad (20)$$

With this operator, we can further calculate the feature matrix of all comprehensive inference source knowledge:

$$S_{k,r} = \left(\nabla_m^c \left((S_k \otimes W_{k,1})^T e_{m,1}, (S_k \otimes W_{k,2})^T e_{m,2}, \dots, (S_k \otimes W_{k,m})^T e_{m,m} \right) \right)^T. \quad (23)$$

In the above equation, \otimes is a comprehensive reasoning simulation operator. If the synthetic reasoning process models human creative thinking as a linear process, then $E_k \otimes W_k = E_k \cdot W_k$, where \cdot is an ordinary matrix multiplication operator. If the comprehensive reasoning process models human creative thinking as a z-order polygon, then \otimes will be defined as follows:

$$\begin{aligned} A_{p \times q} \otimes B_{q \times r} &= C_{p \times r} \Rightarrow c_{i,j} \\ &= \sqrt[z]{\sum_{k=1}^q (a_{i,k} b_{k,j})^z}, \quad i = 1, 2, \dots, p; j = 1, 2, \dots, r. \end{aligned} \quad (24)$$

We can also model creative thinking as a geometric average process, as shown in the equation of (25):

$$c_{i,j} = \sqrt[q]{\prod_{k=1}^q (a_{i,k} b_{k,j})}, \quad i = 1, 2, \dots, p; j = 1, 2, \dots, r. \quad (25)$$

Generally speaking, we can use the comprehensive reasoning model to overload the comprehensive reasoning simulation operator \otimes to achieve the purpose of simulating different types of creative thinking.

Finally, by adding the shape of $E_{k,\text{nor}}$, which is the standard structural element related to the inference result $P_{k,\text{nor}}$ in the process of comprehensive inference, we obtain the parameterized representation $E_{k,r}$: $E_{k,r} = S_{k,r} \otimes \nabla_h^e E_{k,\text{nor}}$ of $P_{k,r}$, where $E_{k,\text{nor}}$ is the matrix form of the parameterized representation of the shape of $P_{k,\text{nor}}$.

The definition of operator \otimes can be overloaded to realize different novel creative thinking activities in intelligent

$$S_k \triangleq E_k - \nabla_n^c(\nabla_h^e E_{k,\text{nor}}), \quad h = \max\{\text{col}(E_{k,l_i}) | i = 1, 2, \dots, n\}. \quad (21)$$

In the above equation, b , E_k , nor are the matrix representations of standard source knowledge at the k -th level.

According to the user input intensity $w_{k,l_i,t}$ for different comprehensive reasoning sources, we can calculate the comprehensive reasoning viewpoint matrix:

$$W_{k,t} \triangleq \left(\nabla_n^c(\omega_{k,l_1,t} \times I_{h \times h}, \omega_{k,l_2,t} \times I_{h \times h}, \dots, \omega_{k,l_n,t} \times I_{h \times h}) \right)^T. \quad (22)$$

Among them, $h = \max\{\text{col}(E_{k,l_i}) | i = 1, 2, \dots, n\} \cdot I_{h \times h}$ is a $h \times h$ dimensional unit matrix.

Now, we can generate a comprehensive reasoning feature result $S_{k,r}$ through the comprehensive reasoning process described by the following form:

systems. Some simple synthetic reasoning simulation operators for topology constructors are as follows:

(1) The arithmetic mean is

$$T_{k,r} = \frac{1}{n} \sum_{i=1}^n (T_{k,l_i} \times \omega_{k,l_i}). \quad (26)$$

(2) The geometric mean is

$$T_{k,r} = \sqrt[n]{\prod_{i=1}^n (T_{k,l_i} \times \omega_{k,l_i})}. \quad (27)$$

(3) The harmonic mean is

$$\frac{1}{T_{k,r}} = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{T_{k,l_i} \times \omega_{k,l_i}} \right). \quad (28)$$

The integrated reasoning process proposed in this paper is essentially either a value process or an extrapolation process. In other words, w^i is the interpolation weight or extrapolation weight of P_{k,l_i}^i , and the constraint $\sum_{i=1}^n w^i = 1$ needs to be satisfied here.

4. Analysis of Poetry Style Based on Text Classification Algorithm

In the research of this paper, the judgment of poetry style based on machine learning, that is, the flow chart of the pattern recognition of poetry text, is shown in Figure 1.

The development and design of the system are shown in Figure 2 according to the research and processing sequence.

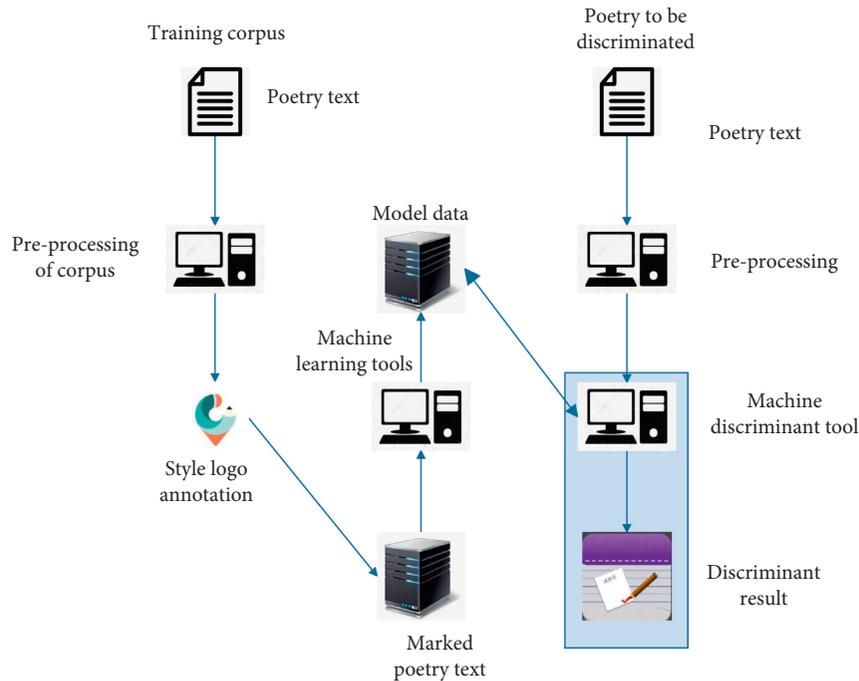


FIGURE 1: Flow chart of pattern recognition of poetry style.

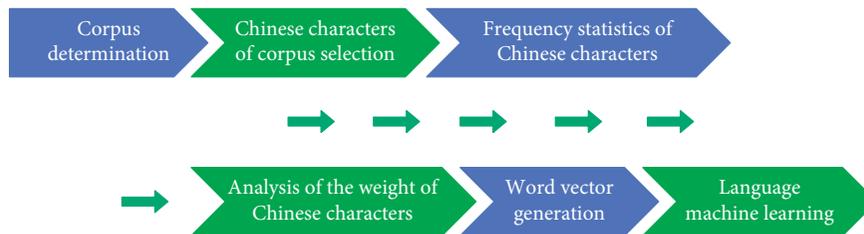


FIGURE 2: The processing flow chart of poem vocabulary material.

Judging the style of poetry is similar to the general process of text classification. First, the algorithm needs to preprocess the poetry documents in the corpus and then map them to the vector space that can be directly processed by the computer. The algorithm selects the poem documents that have been styled in advance as the training corpus and the test corpus. Subsequently, the algorithm uses machine learning methods to generate model data for style classification. Finally, the algorithm builds a machine judging tool for poetry style on the basis of the model data tested on the test corpus. This tool can judge the style of other poetry documents in the corpus. The flow chart of style evaluation is shown in Figure 3.

Poems are composed of text content and belong to unstructured data. It is difficult to analyze them using quantitative analysis methods. It is necessary to introduce data indicators that can be measured to perform data analysis and processing. The commonly used method is to segment the text content and then count the word frequency data. For high-frequency words such as “Wanli,” “Jiangshan,” “Dongfeng,” and “Twilight” that often appear in poems, they can generally express the emotional bias and

style characteristics of the poet’s choice of poetry imagery. In order to classify and aggregate the genres of poets and poems, this paper mainly uses the method of high-frequency word statistics to classify and aggregate the style characteristics of the poets. The cluster analysis of word style is shown in Figure 4.

In order to obtain word vectors with more efficient algorithms, Word2Vec can efficiently train on millions of word lists and hundreds of millions of datasets. The obtained training result-word vector can well measure the semantic relevance and similarity between words. Word2Vec includes two network models: CBOW (Continuous Bag-of-Words) model and Skip-Gram model. The network structure is shown in Figure 5. Skip-Gram can be seen as the reverse process of CBOW. The essence of the CBOW model is a feed forward neural network language model (Feed Forward NNLM).

RNN is a neural network designed specifically for serialization problems, and its basic network structure is shown in Figure 6.

This article summarizes the sequence learning methods of the above two statistical language models. The framework

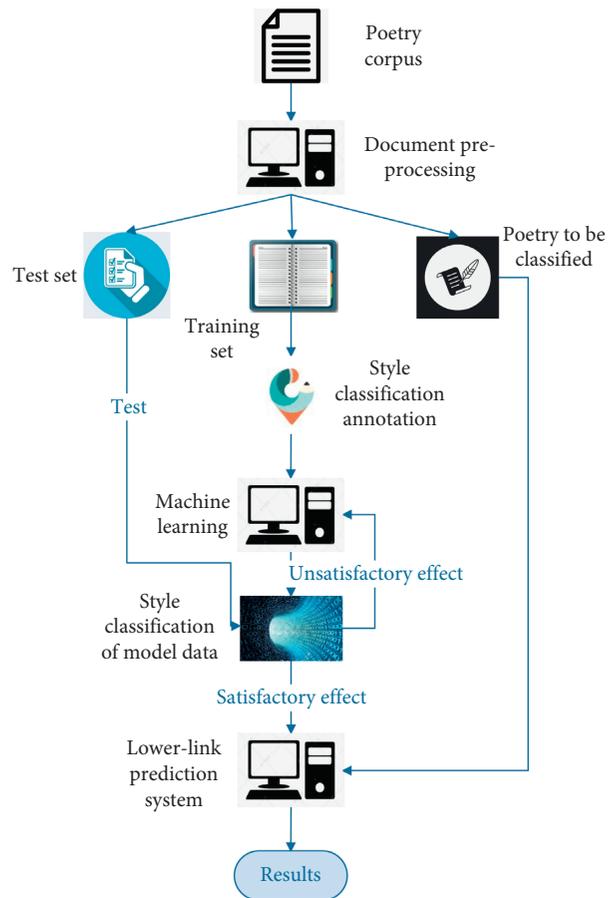


FIGURE 3: Poetry classification process.

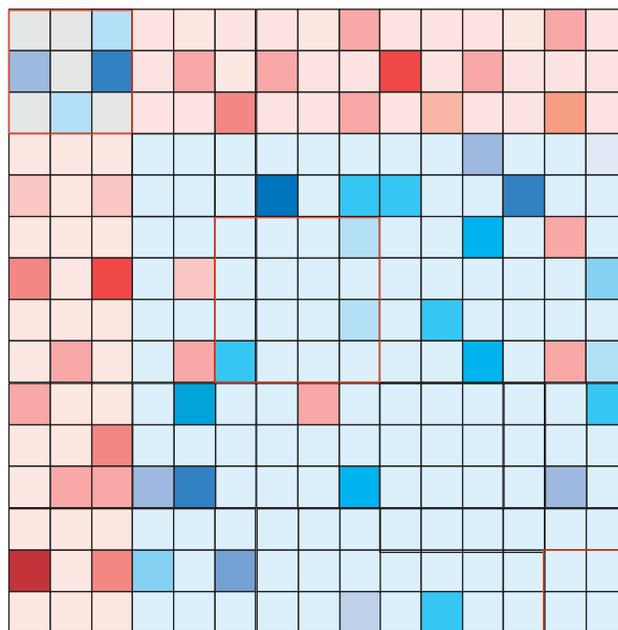


FIGURE 4: Cluster analysis of word style.

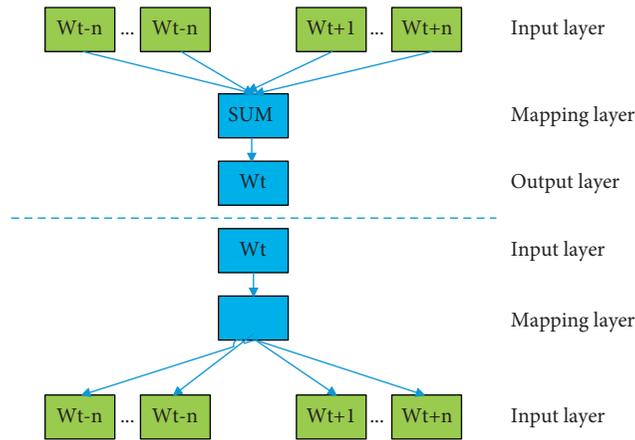


FIGURE 5: Structure diagram of CBOW and Skip-Grām model.

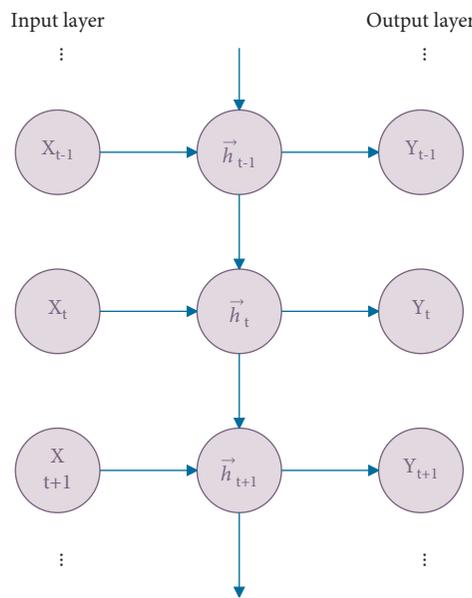


FIGURE 6: Basic network structure diagram.

of the joint language response model based on statistical learning is shown in Figure 7.

In the research of this paper, Brill’s conversion learning method is used to cope with the generated context-sensitive rules in conjunction with the language. The flowchart of the learning algorithm is shown in Figure 8.

The text visualization technology is mainly composed of three modules: text analysis and processing, visual design presentation, and interactive design, as shown in Figure 9. First, the algorithm collects text information, preprocesses data and expresses knowledge, and transforms it into a corresponding structured data format. Then, the algorithm visually presents the information with graphics and color elements through visual coding. In the visual presentation, the algorithm should provide users with effective and reasonable visual presentation and interactive operations based on the user’s cognitive characteristics. At the same time, the

algorithm performs data operations on the visualized information elements through a reasonable interactive method to facilitate users to quickly and clearly understand the displayed information content, thereby realizing a complete visual design implementation process.

After constructing the above model, the effect of the model is verified, and the system function is verified based on actual needs. Moreover, this paper combines the simulation test to verify the effect of the poetry text classification and the poetry style verification of the system of this paper. The poetry text classification is shown in Table 1 and Figure 10.

From the above research, it can be seen that the poetry style analysis method based on the text classification algorithm proposed in this paper has a good classification effect. The experimental results of the poetry style analysis are shown in Table 2 and Figure 11.

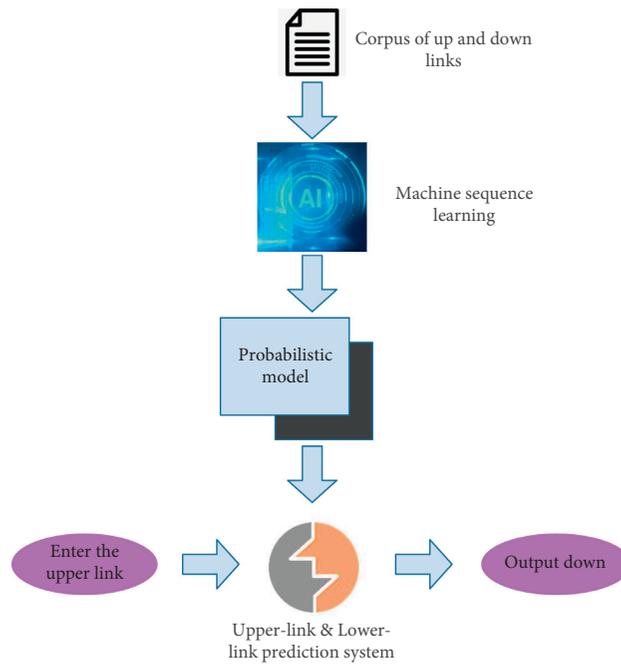


FIGURE 7: Schematic diagram of the joint language response model based on statistical learning.

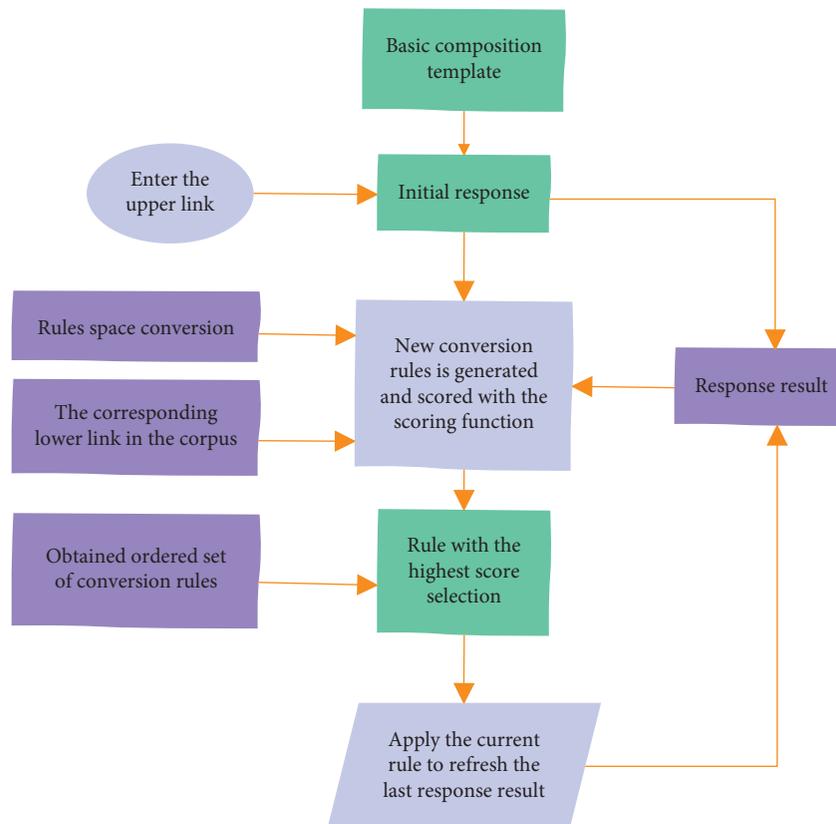


FIGURE 8: Schematic diagram of the learning of the joint language response model based on conversion.

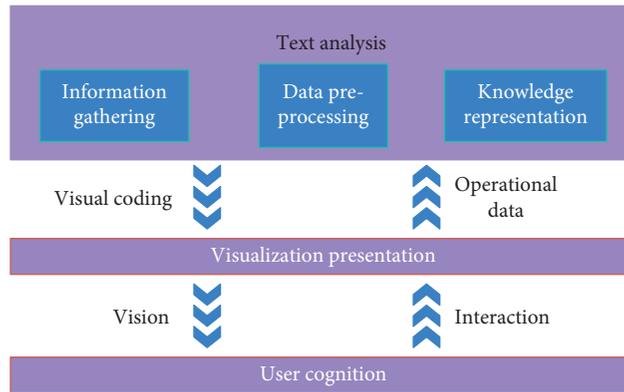


FIGURE 9: Logical architecture diagram of text visualization.

TABLE 1: Experimental data of poem text classification.

NO	Text categorization
1	94.63
2	94.29
3	94.85
4	95.85
5	94.73
6	97.07
7	97.45
8	97.61
9	95.50
10	95.52
11	95.88
12	93.56
13	95.88
14	93.99
15	93.27
16	96.85
17	97.69
18	96.65
19	96.70
20	94.01
21	95.60
22	94.70
23	93.27
24	94.92
25	95.53
26	93.17
27	96.02
28	93.14
29	97.77
30	97.40
31	96.08
32	94.40
33	95.91
34	95.88
35	94.45
36	96.50

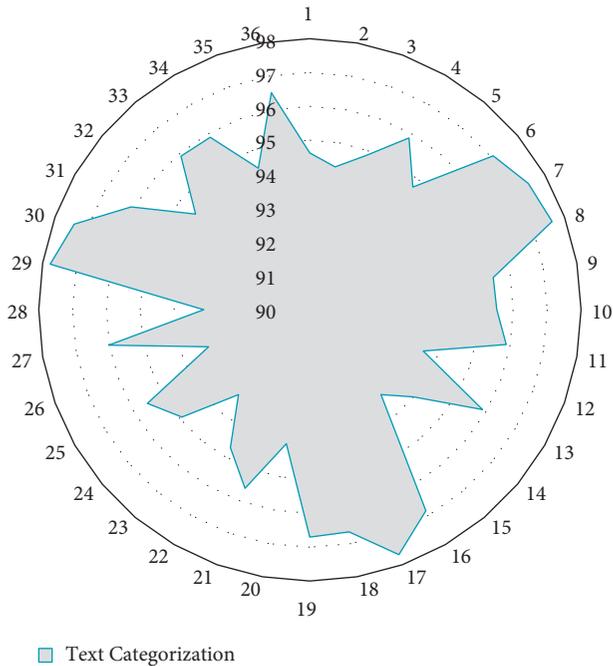


FIGURE 10: Statistical diagram of the evaluation of the classification effect of poetry text.

TABLE 2: Test data of poetry style analysis.

NO	Style analysis
1	75.40
2	76.61
3	76.06
4	88.70
5	82.39
6	84.98
7	91.28
8	83.30
9	81.22
10	78.60
11	83.99
12	81.16
13	79.14
14	84.99
15	84.50
16	79.33
17	73.70
18	74.90
19	89.73
20	87.79
21	76.67
22	89.30
23	91.01
24	76.10
25	89.19
26	80.65
27	90.65
28	74.65
29	91.06
30	90.54
31	84.27
32	84.07
33	77.95

TABLE 2: Continued.

NO	Style analysis
34	89.93
35	90.93
36	87.56

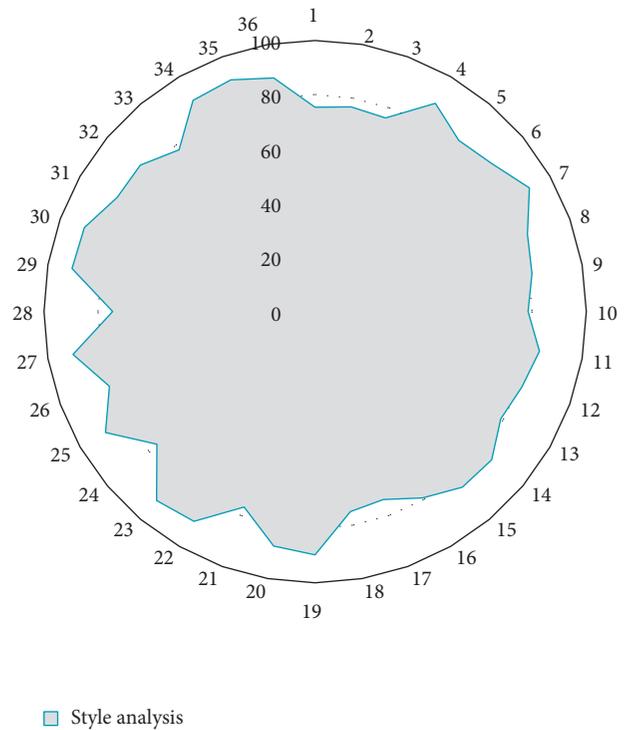


FIGURE 11: Statistical diagram of the effect of poetry style analysis.

From the above research, the effect of the poetry style analysis method based on the text classification algorithm proposed in this paper is very good, which meets the actual needs of poetry style analysis.

5. Conclusion

Before the classification of poetry text begins, the algorithm needs to represent the unstructured document data into a data form that the computer can understand. This requires the algorithm to perform corresponding preprocessing and feature representation of the document first and convert the unstructured or semistructured data form into a structured data form that can be processed by the computer. This includes a series of processes such as word segmentation, filtering of low-frequency words and forbidden words, feature representation, and feature extraction. The so-called word segmentation is to add a separator between each entry in a Chinese document to convert the continuous character stream form of the Chinese document into a discrete word stream form. At present, the main word segmentation methods used are forward and reverse maximum matching method, word-by-word traversal method, best matching method, and word frequency statistics method. In addition,

there are two-scan method, adjacency constraint method, and so on. This paper applies the text classification algorithm to the analysis of the style of poetry and constructs an intelligent classification system. From the research results, it can be seen that the poetry style analysis method based on the text classification algorithm proposed in this paper is very effective and meets the actual needs of poetry style analysis.

Data Availability

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

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This study was sponsored by Henan Industry and Trade Vocational College.

References

- [1] M. H. Aghdam and S. Heidari, "Feature selection using particle swarm optimization IN text categorization," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 5, no. 4, pp. 38–43, 2015.
- [2] S. T. Gissel, "Scaffolding students' independent decoding of unfamiliar text with a prototype of an eBook-feature," *Journal of Information Technology Education: Research*, vol. 14, pp. 439–470, 2015.
- [3] R. R. Kamble and D. V. Kodavade, "Relevance feature search for text mining using FClustering algorithm," *International Journal of Computer Science and Engineering*, vol. 6, no. 7, pp. 223–227, 2018.
- [4] M. Pandi and V. D. K. Rajendran, "Efficient feature extraction for text mining," *Advances in Natural & Applied Sciences*, vol. 10, no. 4, pp. 64–73, 2016.
- [5] M. Mojaveriyan, H. Ebrahimpourkomleh, and S. J. Mousavirad, "IGICA: a hybrid feature selection approach in text categorization," *International Journal of Intelligent Systems Technologies and Applications*, vol. 8, no. 3, pp. 42–47, 2016.
- [6] M. Nayak and A. K. Nayak, "Odia running text recognition using moment-based feature extraction and mean distance classification technique," *Advances in Intelligent Systems & Computing*, vol. 309, no. 3, pp. 497–506, 2015.
- [7] E. Ramalakshmi and S. Golla, "An advanced fuzzy constructing algorithm for feature discovery in text mining," *International Journal of Computers and Applications*, vol. 127, no. 17, pp. 30–34, 2015.
- [8] M. Soleymanpour and H. Marvi, "Text-independent speaker identification based on selection of the most similar feature vectors," *International Journal of Speech Technology*, vol. 20, no. 1, pp. 1–10, 2016.
- [9] T. Oki, "Scene text localization using object detection based on filtered feature channels and crosswise region merging," *Growth & Change*, vol. 21, no. 3, pp. 61–76, 2015.
- [10] A. Tommasel and D. Godoy, "A social-aware online short-text feature selection technique for social media," *Information Fusion*, vol. 40, no. 1, pp. 1–17, 2017.
- [11] G. Wu, M. Zhao, and L. Han, "A fingerprint feature extraction algorithm based on optimal decision for text copy detection," *International Journal of Security & Its Applications*, vol. 10, no. 11, pp. 67–78, 2016.
- [12] Z. Robati, M. Zahedi, and N. F. Far, "Feature selection and reduction for Persian text classification," *International Journal of Computers and Applications*, vol. 109, no. 17, pp. 1–5, 2015.
- [13] T. Zia, Q. Abbas, and M. P. Akhtar, "Evaluation of feature selection approaches for Urdu text categorization," *International Journal of Intelligent Systems Technologies & Applications*, vol. 07, no. 6, pp. 33–40, 2015.
- [14] T. zia, M. P. Akhter, and Q. Abbas, "Comparative study of feature selection approaches for Urdu text categorization," *malaysian journal of computer science*, vol. 28, no. 2, pp. 93–109, 2015.
- [15] Li. De, Z. J. Xue, and C. Lih, "Text recognition algorithm based on text features," *International Journal of Multimedia & Ubiquitous Engineering*, vol. 11, no. 5, pp. 209–220, 2016.
- [16] K. Yan, Z. Li, and C. Zhang, "A New multi-instance multi-label learning approach for image and text classification," *Multimedia Tools and Applications*, vol. 75, no. 13, pp. 7875–7890, 2016.
- [17] G. Kumar and K. Vivekanandan, "Intelligent model view controller based semantic webservice call through mishmash text featuring technique," *Journal of Computational and Theoretical Nanoscience*, vol. 14, no. 4, pp. 2021–2029, 2017.
- [18] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *Ieee Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 11, pp. 2298–2304, 2017.
- [19] M. M. Al-Tahraw, "Polynomial neural networks versus other Arabic text classifiers[J]," *Journal of Software*, vol. 11, no. 5, pp. 418–430, 2016.
- [20] Y. Hu, E. E. Milios, and J. Blustein, "Document clustering with dual supervision through feature reweighting," *Computational Intelligence*, vol. 32, no. 3, pp. 480–513, 2016.