ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

A Path of Constructing Classification System and Semantic Relationships of Yin Shang Oracle Bone Music Rhetoric for Digital Literature Analysis

Jingzhi Lv[1,2*]

[1] School of Music, Anyang Normal University Anyang, 455000, Henan, China
[2] Key Laboratory of Oracle Bone Inscriptions Information Processing, Ministry of Education of China, Anyang Normal University, Anyang, 455000, Henan, China

SUMMARY: *Yinshang oracle bone inscriptions carry the genes of Chinese culture and contain rich musicological information. A classification model based on semantic relationship analysis is proposed for the related document data of Yinshang oracle bone music texts. The model selects document frequency and CHI statistics as the feature selection method, adopts vector space model for text representation, and applies KNN algorithm as the classification algorithm. Through experimental analysis on the dataset, the superior performance of the classification model in this paper is verified, and its average classification accuracy and recall rate reach 89.29% and 8.61%, respectively, which are higher than those of other classification methods with the combination of feature selection and classification algorithms. The classification system of oracle bone music can be categorized into four main types: ritual music, rain prayers, court feasts and court celebrations. In this paper, we refine the classification of oracle bone music rhetoric, which can help to provide practical references of rhetoric categories for oracle bone research, and provide retrieval convenience for related disciplines to utilize the oracle bone materials.*

KEYWORDS: *feature selection; vector space model; KNN; text categorization; Yin Shang oracle bone music rhetoric*

1 Introduction

As one of the eight ancient capitals, Anyang City in Henan Province, bordered by the Zhanghe River in the north and the Taihang River in the west, is the seat of Yinshang culture [1]. The vast majority of oracle bone inscriptions, also known as oracle bone divination, Yinxu script, or tortoise shell and animal bone script, were found in Xiaotun, Anyang City [2, 3]. It is known as the world's four major ancient scripts together with the papyrus script of Ancient Egypt, the clay tablet script of Babylon, and the Mayan script of the American Indians, but the other three scripts have long since disappeared, and only the oracle bone script has developed [4-6]. It has been called the “earliest Chinese character” because it inherited the primitive engraved symbols and bronze inscriptions [7]. Since the discovery of oracle bone inscriptions a hundred years ago, archaeologists have found about 150,000 pieces of oracle bone inscriptions near Xiaotun in Yinxu [8, 9]. Yinshang culture contains a rich content and is an important part of the diverse culture of the Chinese nation [10]. Currently, a

*Lvjingzhi202511@163.com
https://doi.org/10.65102/is2026495
Published: 30 April, 2026

more comprehensive understanding of Yinshang culture can be obtained from the oracle bones, such as hierarchy, agriculture, animal husbandry, and music rhetoric [11].

Among them, music rhetoric, as an important carrier for understanding Yin and Shang music and social hierarchy, beliefs, ritual system, etc., faces new opportunities and challenges in the new period [12]. In traditional oracle bone inscriptions research, the main focus is on their textual interpretation, and the research on their cultural functions and semantic relationships is seriously insufficient, which has become a major flaw in oracle bone inscriptions research [13-15]. Together with the huge number of musical fragments, the traditional manual categorization is inefficient and subjective cognition has a great influence, which hinders the effective research on musical fragments [16, 17]. With the development of information technology, the application of natural language processing and other technologies provides a new path to support the in-depth analysis of related literature, and lays a solid foundation for the extraction of the structural information of music words from the massive oracle bone fragments [18-20]. By constructing the classification system and semantic relationship of Yinshang oracle bone music rhetoric analyzed by digital literature, it can improve the precision of extracting oracle bone information and provide support for the research of Yinshang ritual and music culture [21, 22].

Based on the characteristics of the document data related to Yin Shang oracle-bone lexicon, the study adopts the method of document frequency + CHI statistics for text feature extraction, utilizes the vector space model for machine judgment in text representation, and then selects the KNN text classification algorithm to judge the similarity of the text, and puts forward a classification model for Yin Shang oracle-bone lexicon. Taking the Oracle-Bone Character Document Dataset on (OBI-200) as the dataset, after determining the k-value of the KNN algorithm, different feature selection methods and classification algorithms are selected for testing to check the classification advantage of this paper's model for oracle-bone texts. Then we collect the digital documents of oracle bone lexicon and use the classification model of this paper to discriminate and obtain the classification results of Yin Shang oracle bone lexicon, which will provide the reference of practical literary data for the study of oracle bone literature.

2 Semantic relationship and classification model construction

The practice of assigning a text to the proper categories based on its content using a predetermined classification system is known as text categorization. The numerous texts that need to be categorized serve as the system's inputs, and the categories that correlate to the texts serve as its outputs. The classification model of the Yinshang dynasty's oracle bone music is built using the chosen characteristics and their accompanying meanings in order to create the classification scheme.

2.1 Feature Selection

The so-called feature selection is to select those words that contribute significantly to the classification to represent the text for categorization according to some screening strategy. Commonly used screening strategies are TFIDF, document frequency, information gain, CHI statistics and mutual information.

2.1.1 TFI-DF methodology

The basic principle of TF-IDF is that when a lexical unit appears frequently in one specific class but rarely appears in other classes, it can be regarded as having strong discriminatory power and can therefore be selected for classification. TF-IDF is calculated as:: TF × IDF, where TF refers to lexical frequency, while IDF represents inverse document frequency. TF indicates how often lexical item t appears in document d. The IDF is defined as:

$$IDF = \log \frac{N}{n} \tag{1}$$

where N refers to the total number of documents, and n indicates the number of documents that contain the lemma t.

2.1.2 Documentation frequency

The document frequency (DF) of a lexical entry is the number of documents in the training corpus in which the lexical entry occurs. The use of DF as a feature extraction is based on the following basic assumption: lexical entries with DF values below a certain threshold are low-frequency words, which contain no or less category information. Removing such words from the original feature space not only reduces the dimensionality of the feature space, but also has the potential to improve the classification accuracy.

2.1.3 Information gain

Information gain (IG) is commonly adopted in machine learning. For a lexical item t and a document category c, IG evaluates the occurrence frequency of documents in which t appears or does not appear, so as to calculate the information gain of t for category c. The calculation formula is as follows:

$$\begin{aligned} IG(t) = &-\sum_{j=1}^{m} P\left(c_j\right)\log P\left(c_j\right) + P\left(t\right)\sum_{j=1}^{k} P\left(c_j \middle| t\right) P\left(c_j \middle| t\right) \\ &+P(\bar{t})\sum_{j=1}^{k} P\left(c_j \middle| \bar{t}\right) P\left(c_j \middle| \bar{t}\right) \end{aligned} \tag{2}$$

where $P(c_j)$ denotes the probability that a document in the corpus belongs to class c_j occurs in the corpus, $P(t)$ refers to the proportion of documents containing the lexeme t in the corpus. $P\left(c_j \middle| t\right)$ represents the conditional probability that a document is assigned to class c_j when it contains the lexeme t. $P(\bar{t})$ indicates the probability that a document does not contain the lexeme t and $P\left(c_j \middle| \bar{t}\right)$ denotes the conditional probability that a document belongs to class c_j when the lexeme t is absent. Here, m represents the total number of classes.

2.1.4 CHI statistics

The CHI statistical method is used to quantify the association between a lexical entry t and a document category c. It assumes that the relationship between t and c follows a χ^2 distribution with one degree of freedom. A larger χ^2 value indicates that the word is more closely related to the corresponding category and contains stronger category-specific information. Let N denote the total number of documents in the training corpus, and let c represent a given category. A refers to the number of documents that belong to class c and contain t. B refers to the number of documents that contain t but do not belong to class c. C indicates the number of documents that belong to class c but do not contain t. D represents the number of documents that neither belong to category c nor contain t. Therefore, the CHI value of t for c is computed as:

$$\chi^2(t,c) = \frac{N \times (AD - CB)^2}{(A+C)(B+D)(A+B)+(C+D)} \tag{3}$$

For multi-category classification, the CHI value of t is first calculated for each category independently. Then, the overall CHI value of the lexeme t in the whole corpus is obtained through the following formula:

$$\chi^2_{\max}(t,c) = \max_{i=1}^{m} \chi^2(t,c_i) \tag{4}$$

where m denotes the total number of categories.

2.1.5 Mutual information

Mutual information (MI) is frequently applied in statistical language modeling. In this method, A represents the number of documents that contain the lexeme t and belong to category c, B indicates the number of documents that include t but are not assigned to c, C refers to the number of documents that fall into category c without containing t, while N denotes the total number of documents in the corpus. Based on these definitions, the mutual information between t and c can be calculated as follows:

$$MI(t,c) \approx \log \frac{A \times N}{(A+C) \times (A+B)} \tag{5}$$

If t and c are statistically independent, namely (i.e., $P(tc) = P(t) \times P(c)$, the value of $I(t,c)$ is naturally zero. To extend mutual information to multi-category tasks, its calculation is handled in a way similar to CHI statistics. The mutual information of t for category c is obtained through the following equation:

$$MI_{\max}(t) = \max_{i=1}^{m} I(t,c_i) \tag{6}$$

where m refers to the number of categories. Entries below a preset threshold are deleted from the initial feature space, thereby reducing the feature dimension, while entries above the threshold are retained.

In this paper, we adopt the method of DF + CHI for feature extraction, i.e., we first use DF to remove low-frequency entries below a certain threshold to eliminate the reliance of CHI on low-frequency entries. Then CHI is used to remove the noisy words with low category information from the remaining words, so that the words are evenly distributed in each category.

2.2 Text representation

The vector space model (VSM) provides an effective representation method for large-scale corpus categorization.

(1) Vector Space Model (VSM): Given a natural language text $D = D(t_1, w_1; t_2, w_2 \cdots t_n, w_n)$, the recurrence relation and sequence order of t_k in the text make direct analysis relatively difficult. To simplify the analytical process, the order of t_k can be disregarded, and each t_k is assumed to be mutually exclusive, meaning that no repetition is considered. In this case,r $t_1, t_2, \cdots t_n$ can be treated as an n-dimensional coordinate system, while $w_1, w_2, \cdots w_n$ serve as the corresponding coordinate values. Thus, a text can be expressed as a vector in an n-dimensional space, which is defined in this study as: $D = D(w_1, w_2, \cdots w_n)$. This representation is also called the vector representation or vector space model of text D.

(2) Similarity measure: The correlation degree between two texts D_1 and D_2 is measured by their similarity $Sim(D_1, D_2)$. Under the vector space model, text similarity can be determined by calculating the distance between the corresponding vectors.

Similarity is often calculated by the inner product between vectors: $Sim(D_1, D_2) = \sum_{k=1}^{n} w_{1k} * w_{2k}$.

Alternatively, it is expressed in terms of the angle cosine: $Sim(D_1, D_2) = \cos\theta = \frac{\sum_{k=1}^{n} w_{1k} * w_{2k}}{\sqrt{\left(\sum_{k=1}^{n} w_{1k}^2\right) * \left(\sum_{k=1}^{n} w_{2k}^2\right)}}$.

In this paper, vector space model (VSM) is used to represent the text of Yin Shang oracle bone music, and combined with the related digital literature corpus, a text classification model is established to classify the Yin Shang oracle bone music for machine judging. The main formulas are as follows:

$$\bar{X} = (w_1, w_2, \cdots w_n) \tag{7}$$

$$Sim(\bar{X}, \bar{Y}) = \frac{\bar{X} \cdot \bar{Y}}{\|\bar{X}\| \times \|\bar{Y}\|} \tag{8}$$

$$TFIDF_i = TF_i \times \log(N / DF_i), i = 1, 2 \ldots n \tag{9}$$

$$w_i = \frac{TFIDF_i}{\sqrt{\sum_{j=1}^{n} (TFIDF_j)^2}}, i = 1, 2 \ldots n \tag{10}$$

where $Sim(\bar{X}, \bar{Y})$ denotes the model's evaluation value of the similarity of the Yin Shang Oracle Bone Lexicon text, and $\bar{X}, \bar{Y}$ denotes the instance feature vectors corresponding to the model, where each component w_i denotes the canonicalization weights of the corresponding feature in the text. Here, this paper adopts the TFIDF method to calculate the weights, $TFIDF_i$ denotes the TFIDF value of the lexeme t_i in the text, TF_i denotes the frequency of occurrence of the lexeme t_i in the text of the Yin Shang Oracle Bone Music Rhetoric, N denotes the total number of all the training documents, and DF_i denotes the frequency of the document that contains the lexeme t_i. In order to reduce the over-suppression of high-frequency features to low-frequency features, $TFIDF_i$ is normalized in the experiment, and w_i is finally obtained by normalization calculation.

2.3 KNN classification method

K-Nearest Neighbor (KNN) is one of the best classification algorithms under the Vector Space Model (VSM), which is simple but stable and accurate, and has been widely used in the fields of text categorization, face recognition, medical image processing, and text recognition, etc. The core idea of the KNN algorithm for text categorization is to compute the similarity between the test text to be tested and each text in the training sample set, to find the K texts with the highest similarity, calculate the similarity weight of each category, and attribute the test text to the category with the largest similarity weight. In this paper, KNN algorithm is used as the classification method of text classification model.

The specific principle and process are as follows, Yin Shang oracle bone music text $D_i=\left(t_{i1},w_{i1};t_{i2},w_{i2};......t_{in},w_{in}\right)$, t_{ik} and w_{ik} represent the k feature word and its weight value of the Yin Shang oracle bone music text, respectively, $V_i=\left\{w_{i1},w_{i2},......,w_{in}\right\}$ represents the text feature vector of Yin Shang oracle bone music text D_i. Suppose the training set of Yin Shang oracle bone music text is $S=\left\{S_1,S_2,......S_m\right\}$, m denotes the number of Yin Shang oracle bone music categories in the training set, where $S_i=\left\{V_{i1},V_{i2},......V_{iz}\right\}$, represents the ith class in S, i.e., the z thYinshang oracle bone music text vector in S_i, $i\in[1,m]$, and V_{iz} represents the textual feature vector of the z thYinshang oracle bone music text in the ith class. In order to discriminate the categories of Yin Shang oracle bone music rhetorical texts:

(1) First, calculate the similarity between v_i and all the Yin Shang oracle bone music texts. The similarity can be calculated using Euclidean distance or cosine similarity. Assume that given two Yin Shang oracle bone music texts with D_j, the feature vectors of their Yin Shang oracle bone music texts can be expressed as $V_i=\left\{w_{i1},w_{i2},......,w_{in}\right\}$ with $V_j=\left\{w_{j1},w_{j2},......,w_{jn}\right\}$, n is the feature dimension of the text vector of the Yin Shang oracle bone music dictionaries, and the Euclidean distance between the texts of the Yin Shang oracle bone music dictionaries is shown below:

$$Dis\tan ce\left(D_i,D_j\right)=\sqrt{\left(w_{i1}-w_{j1}\right)^2+\left(w_{i2}-w_{j2}\right)^2+\cdots+\left(w_{in}-w_{jn}\right)^2} \tag{11}$$

The cosine similarity formula between the texts of the Yin and Shang oracle bone music dictionaries is shown below:

$$Sim\left(D_i, D_j\right)=\frac{\sum_{k=1}^{n} w_{ik} w_{jk}}{\sqrt{\sum_{k=1}^{n} w_{ik}^2} \sqrt{\sum_{k=1}^{n} w_{jk}^2}} \tag{12}$$

(2) Sort them according to the similarity $Sim\left(D_i, D_j\right)$ from the largest to the smallest.

(3) Select K most similar text samples of Yin Shang oracle bone music rhetorics as the best nearest neighbors of the test document D_i.

(4) Calculate the affiliation degree of Yin Shang oracle bone music rhetoric text D_i in each Yin Shang oracle bone music rhetoric category based on the K nearest neighbors, as shown in Equation (13):

$$p\left(D_i, S_m\right)=\sum sim\left(D_i, D_j\right) \delta\left(D_j, S_m\right) \tag{13}$$

where: $\delta\left(D_j, S_m\right)$ denotes that if the text of Yinshang oracle bone music rhetoric D_j belongs to the category of hidden dangers S_m, the value is 1, otherwise the value is 0, and the formula is shown in Equation (14):

$$\delta\left(D_j, S_m\right)=\begin{cases}1, & D_j \in S_m \\ 0, & D_j \notin S_m\end{cases} \tag{14}$$

(5) Select the Yin Shang oracle bone music type S_m with the largest degree of affiliation, and assign the test oracle bone music text D to that oracle bone music type S_m.

3 Experimentation and analysis

3.1 Data sets

In this section, the effectiveness of the proposed method will be tested on the Oracles-Oracle Bone Character Document Dataset (OBI -200).The OBI -200 dataset is a randomly selected set of 200 oracles from the Oracle Bone Collocation Collection, which includes 30 pairs of oracles with the same text, and by splitting and removing deactivated words, we finally obtain the 200 × 386 oracles- oracle bone character matrix, which will be processed by DF + CHI and then tested. Among them, the oracle bone divination is mainly categorized into three types: chaste divination, sacrifice, and hunting patrol, and the experimental data are divided into training data and test data according to 8:2 of the total sample data.

3.2 k-value determination

The constructed classification model of Yin Shang Oracle Bone Le Rhetoric uses the K nearest neighbor algorithm as the classification method, while too small a k value of the KNN algorithm will lead to the effect of outliers being magnified, resulting in too high a variance, in other words, a smaller k value means that the overall model becomes complex and is prone to overfitting. Whereas too large a k value will result in low reference value samples at a distance being over considered and the bias increasing, in other words, a larger k value means that the overall model becomes simpler and is prone to underfitting, as well as increasing the computational overhead. Therefore an appropriate k value needs to be determined experimentally. The OBI -200 dataset was utilized for testing, and the change in correct rate and change in elapsed time for different values of k are shown in Figure 1, respectively.The KNN classifier determines the parameter value of k=2 based on the correct rate and elapsed time results, at which time the maximum correct rate is 0.897, and the minimum elapsed time result is 614 s. The KNN classifier is also used to determine the correct rate and the elapsed time result.

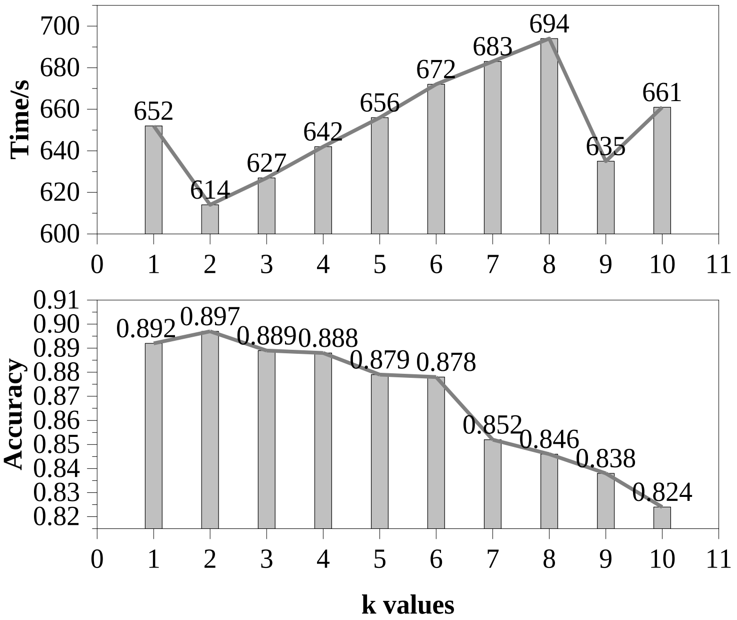

Figure 1: Changes in the accuracy and time of different k values

3.3 Model classification effect

In this paper, accuracy (P) and recall (R) are adopted as evaluation indicators to assess the performance of the KNN-based classification model.

To verify the effectiveness of the proposed mood classification model for classical poetry classification, comparative experiments are carried out using different feature term sets. These include TF-IDF, IG+MI, and the proposed DF+CHI feature selection method. Meanwhile, three classifiers are selected for comparison, namely KNN, Support Vector Machine, and Naive Bayes. The accuracy and recall results of different classification models are presented in Figures 2 and 3.

The results show that, under the same classifier, the model using DF+CHI as the feature selection method achieves higher accuracy and recall than those using TF-IDF and IG+MI. From the perspective of classifier comparison, the performance difference between Naive Bayes and Support Vector Machine is relatively small. However, both methods obtain lower accuracy and recall than the K-Nearest Neighbor classifier. When the classifier is KNN and the feature selection method is DF + CHI, which is the classification model constructed in this paper, at this time, the model has the highest accuracy and recall curves (in blue), and the classification accuracy for the three types of themes in the divination-Oracle Bone Characters document dataset, namely Chaste Divination, Sacrifices, and Hunting Patrols, is 89.15%, 90.04%, and 88.69%, with an average accuracy rate of 89.29%, and the recall rates are 88.37%, 90.26%, and 87.20%, with an average recall rate of 88.61%. Therefore, the classification model of Yin Shang oracle bone lexicon in this paper has achieved relatively satisfactory results in the experiments.

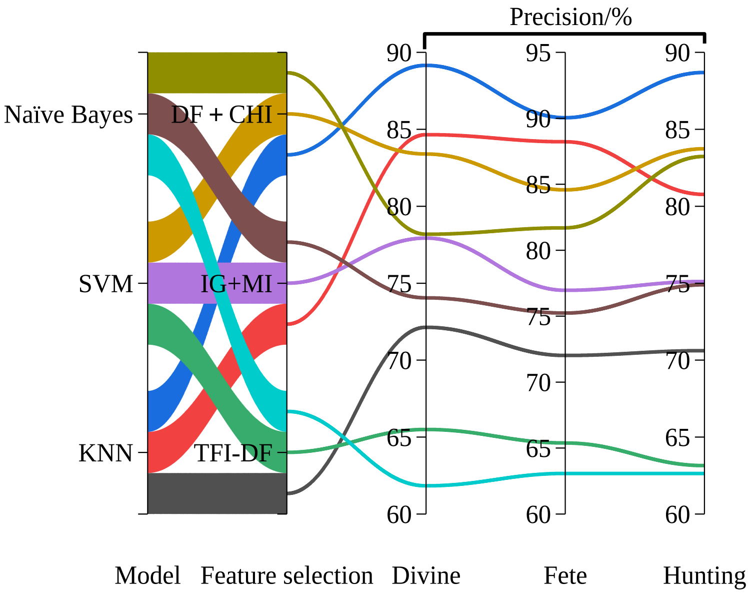

Figure 2: Precision value of different classification models

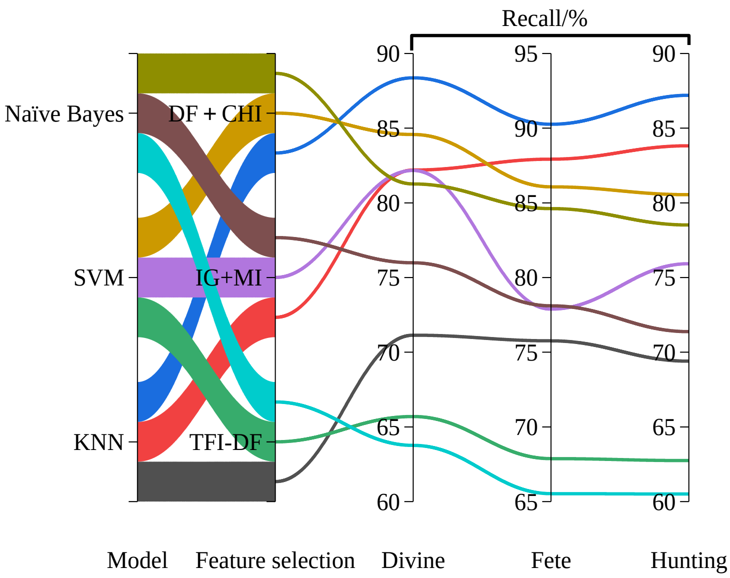

Figure 3: Recall value of different classification models

3.4 Discussion of thresholds

In the selection of feature items, it will involve the determination of the value of the threshold of the number of feature items, and this paper determines its threshold based on the average accuracy of classification, combined with the overhead of time.

Generally, the optimal threshold varies with different corpora. Therefore, this paper explores the changing pattern of threshold values through experiments and determines an approximate effective range. First, the execution time overhead is mainly affected by the number of feature items. Thus, different feature-item quantities are selected for time-cost statistics, and the results are displayed in Figure 4. As the number of feature items increases, the time overhead also rises. When the number of feature items is around 100, the execution time is approximately 40 seconds. When the number increases to nearly 220, the time exceeds 103 seconds, showing a rapid growth trend.

The accuracy under different numbers of feature items is shown in Figure 5. Although Naive Bayes, SVM, and KNN are different classification models, their overall accuracy tends to improve as the number of feature items increases. However, when the number of feature items exceeds 100, the accuracy change gradually becomes stable. Considering both time overhead and classification accuracy, the number of feature items used in this paper is finally set between 100 and 150.

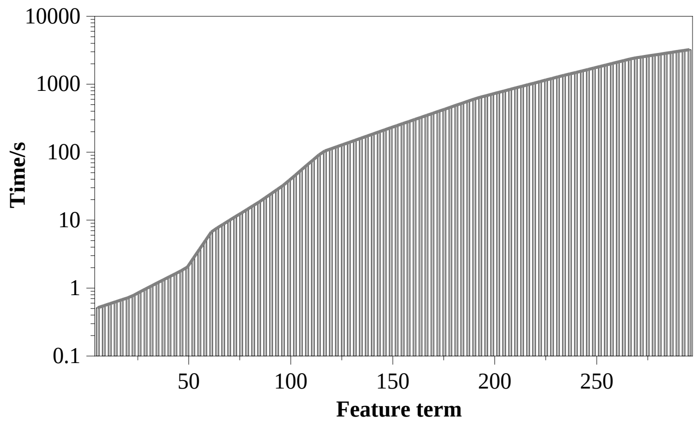

Figure 4: The time cost of different eigenitems amount

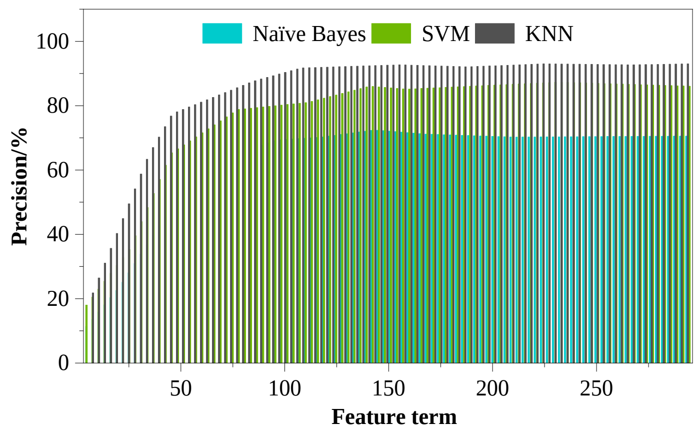

Figure 5: The precision value of different eigenitems amount

4 Classification System of Yin and Shang Oracle Bone Music Rhetoric

4.1 Yin and Shang Oracle Bone Music Rituals

The Yin and Shang oracle-bone inscriptions hold the mysteries of Chinese culture and demonstrate the social conditions of the time when the characters were created. The Chinese characters themselves provide a glimpse of the unique cultural psychology of the ancients, of cultural elements such as tools, dwellings, institutions, languages, philosophies, beliefs, customs, and behavioral norms, and of their evolution. From content to form, Chinese characters mark the vivid existence of Chinese culture. The oracle bones, with their many names of musical instruments, are a precious source of ancient music.

In the social life of the Yin and Shang, music, song and dance were prevalent, and the subjects of their contents were derived from the expression of people's feelings during their survival activities such as production practices, life pursuits, hero worship, folk customs, war scenes or religious beliefs. Such as the “Poetry - Ode to Commerce - that” in the specific depiction of the Shang Dynasty ancestor act, up with a variety of names and types of musical instruments, in the drums, tubes, chimes, yongs, the sound of the dance team in full bloom, and music, buckle the rhythm, singing the hymns, boards and orderly, and finally in the feast to the end.Instrumental music, singing and dancing are integrated into one. In the oracle bone inscriptions of Yinxu, there is a record of a batch of names for music and dance songs, among which "Shangzou", "Zoushang", "Xueshang", "Meizhou", "Gezou", "Jiazou", etc., all refer to different names for sacrificial music or songs.

In short, the Yinshang society music song and dance is very prosperous, many types of musical instruments, Yin ruins oracle bones and stone inscriptions seen, not only a series of musical instruments, there are a number of names of Shang, beauty, each, Jia, new, old, Chi, a number of different sacrificial songs. Music and dance in the Yinshang aristocratic social life is very prosperous, the ruler of the ritual music order government, different music and dance for different occasions, to reflect the majesty, harmonize the upper and lower, to raise the dignity of its quality.

4.2 Oracle Classification

Using the constructed KNN classification model to classify Yin Shang oracle bone music, we first collect the related digital literature of Yin Shang oracle bone texts, and then apply the trained classification model to generate the classification results of Yin Shang oracle bone music, which are shown in Fig. 6. The colors represent the different themes of Yin and Shang oracle bone music, and the theme class represented by the green node is the sacrificial music, such as the “Zuo Shang” related engraved words in the ancestor worship.The blue nodes represent rain-invoking songs, such as "Each performs a piece, then again, there is heavy rain," with the sacrifices directed towards rain gods, river gods, and the like.The gray nodes represent the court banquets and drinks, and the red nodes represent the court celebrations, such as the engraving of “playing in the new room”, which is a musical piece played in the celebration of the new palace. It shows that the proposed KNN classification model is able to differentiate the oracle bone music words of different theme categories.

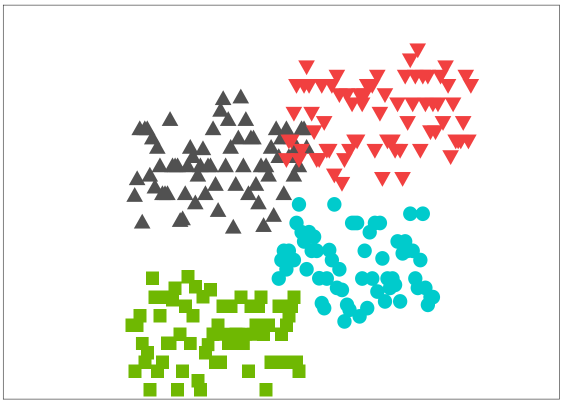

Figure 6: The classification results of the Yinshang oracle bones music word

5 Conclusion

The purpose of this paper is to provide an efficient automated solution for Yin Shang oracle bone music speech classification with the help of current mature natural language processing technology. Based on the semantic relation representation method and the K-nearest neighbor algorithm, a classification model of Yin Shang oracle bone music is constructed. The k-value parameter at the maximum correct rate (0.897) and the minimum time spent (614s) is selected, and the parameter value is determined as 2. The classification model with different feature selections and different combinations of classifiers is analyzed on the Buqi-Oracle Bone Character Document dataset, and the results show that the model in this paper has a higher classification accuracy and recall than the other approaches, and its average accuracy rate of classifying three types of topics: chaste divinations, sacrificial offerings, and hunting patrols, is higher than the other approaches. The average accuracy and average recall of the three types of topic classification reaches 89.29% and 8.61%, which improves the classification effect to a certain extent, indicating that the classification model of this paper, which uses DF + CHI as the feature selection method, the vector space model as the text representation method and the KNN algorithm as the classifier, has a good performance of oracle bone corpus classification. For the digital documents of Yinshang oracle bone music speech, after applying the model for classification, four oracle bone music speech types are obtained, which are ritual music speech, praying for rain, court feast and drinking music speech, and court celebration music speech. In summary, the proposed classification model can effectively identify the semantic relationships among the texts of Yin Shang oracle bone music and realize the classification of Yin Shang oracle bone music.

Although the proposed classification method has achieved good performance in analyzing the text types of oracle bone music, there are still some areas that need to be improved: the data size of the oracle bone music texts is small, and the data set size needs to be enlarged continuously in the subsequent research. We introduce Oracle domain knowledge, such as examples, phrases, grammar, etc., and construct a semantic similarity network based on domain knowledge to continuously improve the model performance.

About the Author

Jingzhi Lv was born in Foshan, Guangdong, P.R. China, in 1981. He obtained a master's degree from Xi'an Conservatory of Music in China. I am currently working in Anyang Normal University. My main research direction is music and oracle bone inscriptions in Yin Ruins.

References

[1] Guangkuo, Y. (2013). The discovery and study of the early Shang culture. A companion to Chinese archaeology, 323-342.

[2] Da WEIH, D. (2023). Strategies of Graphing Ideas in Shang Oracle Bone Scripts. The International Journal of Chinese Character Studies, 6(2), 175-213.

[3] Li, C., Ding, Z., Hu, X., Li, B., Luo, D., Wu, A., ... & Ji, R. (2025). Oraclefusion: Assisting the decipherment of oracle bone script with structurally constrained semantic typography. In Proceedings of the IEEE/CVF International Conference on Computer

Vision (pp. 19893-19902).

[4] Autran, P. O., Dejoie, C., Dugand, C., Gervason, M., Bordet, P., Hodeau, J. L., ... & Martinetto, P. (2023). Illustrating papyrus in ancient Egypt. Scientific Reports, 13(1), 524.

[5] Robson, E. (2019). The clay tablet book in Sumer, Assyria, and Babylonia. A Companion to the History of the Book, 173-190.

[6] Law, D., & Stuart, D. (2017). Classic Mayan: An overview of language in ancient hieroglyphic script. The Mayan Languages, 128-172.

[7] Chen, T., Qian, Y., Pei, J., Wu, S., Wu, J., Li, L., & Tu, J. Y. (2020). A study on encoding-based oracle bone script recognition. Journal of Chinese Writing Systems, 4(4), 281-290.

[8] Anzhu, L. (2020). Oracle-bone inscriptions and cultural memory. Front. Art. Res, 2, 63-73.

[9] Yabing, S. (2020). Manifold and splendid: 120 Years of research on the oracle bone inscriptions and Shang history. Chinese Studies in History, 53(4), 351-368.

[10] Ning, B. (2025). Study on the visual effect and cultural symbol of Yin Shang bronze. Mediterranean Archaeology & Archaeometry, 25(1).

[11] Li, C., Li, S., Guo, C., & Liu, S. (2025). An Overview of Ritual Culture in Shang and Zhou Dynasties. The Chinese History of Literary Theory in Pre-Qin Period, 109-130.

[12] Zhaoming, Z., Punvaratorn, M., &Phansue, V. (2025). The Development of Pre-Qin Music in the Zhong Yuan of China. Journal of Multidisciplinary in Humanities and Social Sciences, 8(5), 2020-2042.

[13] Wang, M., Cai, Y., Gao, L., Feng, R., Jiao, Q., Ma, X., & Jia, Y. (2022). Study on the evolution of Chinese characters based on few-shot learning: From oracle bone inscriptions to regular script. Plos one, 17(8), e0272974.

[14] Gao, F., Chen, X., Li, B., Liu, Y., Jiang, R., & Han, Y. (2024). Linking unknown characters via oracle bone inscriptions retrieval. Multimedia Systems, 30(3), 125.

[15] Kim, H. (2023). A study on the relationship between the period of oracle bone inscriptions and the development of character forms. Journal of Chinese Writing Systems, 7(2), 86-89.

[16] Song, S., & Yang, Z. (2025). Beyond “human-deities dialogue”: a study of oracle bone divination in the Shang dynasty from the perspective of human-deities communication. International Communication of Chinese Culture, 1-22.

[17] Wang, P., Zhang, K., Wang, X., Han, S., Liu, Y., Wan, J., ... & Liu, Y. (2024). An open dataset for oracle bone character recognition and decipherment. Scientific Data, 11(1), 976.

[18] Fu, X., Yang, Z., Zeng, Z., Zhang, Y., & Zhou, Q. (2022). Improvement of oracle bone inscription recognition accuracy: A deep learning perspective. ISPRS International Journal of Geo-Information, 11(1), 45.

[19] Han, X., Bai, Y., Qiu, K., Liu, Z., & Sun, M. (2020, October). Isobs: An information system for oracle bone script. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 227-233).

[20] Liu, M., Liu, G., Liu, Y., & Jiao, Q. (2020). Oracle bone inscriptions recognition based on deep convolutional neural network. Journal of image and graphics, 8(4), 114-119.

[21] Wang, M., & Deng, W. (2024). A dataset of oracle characters for benchmarking machine learning algorithms. Scientific Data, 11(1), 87.

[22] Zhou, S., Wang, X., Qiu, J., Bu, W., & Wang, H. (2025). OracleNet: enhancing Oracle Bone Script recognition with Adaptive Deformation and Texture-Structure Decoupling. npj Heritage Science, 13(1), 273.