

# Research on the Classification Algorithms for the Classical Poetry Artistic Conception based on Feature Clustering Methodology

Jin-feng LIANG<sup>1, a</sup>

<sup>1</sup> Guangxi Technical College of Machinery and electricity, Nanning 530007, China

<sup>a</sup> liangjinfeng@126.cm

**Keywords:** Classification Algorithm; Classical Poetry Artistic Conception; Fuzzy Clustering.

**Abstract.** From ancient to nowadays, myriad literati and poets had wrote many popular works, these works is the pearl of Chinese literature and treasures of traditional culture. Due to the length limit of classical poetry, there often appear a lot of allusions. In this paper, we conduct theoretical analysis and numerical simulation on the classification algorithms for the classical poetry artistic conception based on feature clustering methodology. Artistic conception of Chinese classical poetry is an important aesthetic category in the theory which is also an important topic of Chinese traditional aesthetics. It is necessary to improve the classical text classification model and then applied to the classification of Chinese classical poetry. By a large number of classical poetry can be found after reading and classical Chinese poems are often not directly ventilation thought emotion. Our proposed method performs well in the experimental section.

## Introduction

China has the value reputation and notation of poetry kingdom. From ancient to nowadays, myriad literati and poets had wrote many popular works, these works is the pearl of Chinese literature and treasures of traditional culture. Due to the length limit of classical poetry, there often appear a lot of allusions, lingo in his works and the content is often difficult to understand for ordinary readers so for a long time, the study of classical poetry is often some experts and scholars in this field has deep attainments. These experts and scholars often knowledgeable will appear in the poetry of some specific content has the keen insight. For automatic classification system, its main characteristic is based on statistical method and based on machine learning method is applied to automatic text classification. Among them and based on statistical method is carried out on the characteristics of the term in the document statistics through the statistical results to analyze the content of the document. The method based on machine learning is to use the characteristics of the computing speed is fast to a large number of statistical content through the computer to find the connection between them and thus obtains the conclusion of some ordinary people is not so easy to find. So will be based on the statistical method combined with machine learning to deal with the text for classification is a good way. Based on statistical classification method is, in essence, a quantitative reasoning under uncertainty. It through the statistics and analysis on the large-scale corpus information needed to gain knowledge so that they can on the whole document of after data processing provides more objective data quality with high coverage and good consistency. Of course it may ignore the smaller as a proportion of the document category. For algorithm based on connected it should be able to improve the parallelism of computation and the realization of the distribution of information access advantage and also has high fault tolerance for we have to learn the complex nonlinear mapping is a good choice.

Therefore, to deal with the mentioned challenges and drawbacks, we conduct theoretical analysis and experimental implementation of the classification algorithms for the classical poetry artistic conception based on feature clustering methodology. Through sort of classical poetry into the problem of text categorization we can make use of some classical machine learning method, compared with general text but Chinese classical poetry has a special characteristic especially compared with the modern text difference is bigger. Therefore it is necessary to improve the classical

text classification model and then applied to the classification of Chinese classical poetry. By a large number of classical poetry can be found after reading, classical Chinese poems are often not directly ventilation thought emotion but by a word or words in the poem readers to depict the picture and the song poetry through these pictures connect. When readers after reading the whole works like the movies and these pictures show readers dynamically so as to bring readers a kind of artistic conception, it is through this kind of artistic conception that readers feel want to express the content of the work. In the following sections we will discuss the algorithm in detail [1-4].

## Our Designed Algorithm and Approach

**The Concepts of Clustering and Classification.** As we say, data annotation is an additional, expensive, and error-prone preparation process. Individual data have to be carefully inspected in order to pinpoint somewhat reliable class labels for the training patterns. Instances of the difficulties involved in the process are found in areas such as bioinformatics, speech processing, or affective computing, where the exact class labels may not even be explicitly observable. In the following formula one, we define the data set for further processing.

$$S = \{(x_i, y_i) | x_i \in R^d, y_i \in Y, i = 1, 2, \dots, m\} \quad (1)$$

Entry document frequency refers to the corpus in the terms of the number of the document. Document frequency is below a certain threshold of entry is low frequency words and will be removed from the original feature space this entry will not only reduce the dimension of feature space and could improve the accuracy of classification. Mutual information is widely used in statistical language model which is shown below in the formula 2.

$$MI(t, c) = \log_2 \left( \frac{p(t \wedge c)}{p(t) \times p(c)} \right) \quad (2)$$

Multi-label learning is a kind of complicated decision task and the same object can belong to multiple categories. At the same time such tasks in text classification, image recognition and gene analysis in areas such as widespread. Multi-label classification task often described by the high-dimensional feature and there are a lot of irrelevant and redundant information. Currently has a large number of single tag feature selection algorithm is proposed to cope with the problem of dimension disaster but for many tag attributes reduction and feature selection are rarely studied. Applying rough set to multiple tags feature selection and the data for multi-label classification task which will be redefined the neighborhood rough set calculation method of the lower approximation and dependence shown in the formula 3 and the table one.

$$U = \{x_1, x_2, \dots, x_n\}; A = \{a_1, a_2, \dots, a_n\}; D = \{l_1, l_2, \dots, l_n\} \quad (3)$$

Table.1 The Example of Single-Label Task for Clustering

| $U$   | $a_1$ | $a_2$ | $a_3$ | $D$   |
|-------|-------|-------|-------|-------|
| $x_1$ | -1    | 0     | 0     | $l_1$ |
| $x_2$ | 1     | 0     | 1     | $l_2$ |
| $x_3$ | 2     | 1     | 1     | $l_3$ |
| $x_4$ | 3     | 2     | -1    | $l_4$ |

From the perspective of classification, pattern recognition is one kind of the specific things to specific process. That is to use a certain number of samples, classifier design according to the similarity between them and then use the designed classifier to identify the samples classification decisions. Classification process can be either in the original data space and also can to transform the original data the data image to the most can reflect the nature of classification in the feature space. In contrast, the latter makes it easier to decision the design of the machine and it said through a more stable characteristics which will improve the decision-making performance of the machine and

remove redundant or irrelevant information and more easy to find the inherent relationship between the object of study. The following expression shows the process.

$$f(C, t) = g(C) - p(t)g(C|t) + p(\bar{t})g(C|\bar{t}) \quad (4)$$

Another key task of pattern recognition is for data analysis and processing. In general, the pattern recognition system described by the characteristics and attributes contained in the object of the mathematical model of object of study in general depends on the mutual relations between characteristics. In cognitive science, the relationship between objects is known a lot of situations and the cause of significant differences between the objects are unknown.

**The Principle of Classical Poetry Artistic Conception.** Of classical Chinese poetry in the high artistic achievement, the roll degree we should to cherish today and look at the ancients in poetry creation experience and aesthetic. Artistic conception of Chinese classical poetry is an important aesthetic category in the theory which is also an important topic of Chinese traditional aesthetics. We from the creative practice of classical poetry and contact the ancient literary theory which can be in the vast world to sum up the experiences of ancient poets to create artistic conception of art and explore classical poetry show the art law of artistic conception. In the objective world, the object is specific and limited, but the poet through the selection of image and an orderly combination of image creation of artistic conception is a visionary blurred, is endless. Poet in choosing objective image build stereo space artistic conception of classical poetry, because of the limitations of objective image and artistic conception in the works of classical poetry infinite extensibility, classical poetry artistic conception is bound to produce many don't need to use a specific object cannot use to fill the blank of specific image, thus forming the classical poetry artistic conception of the actual situation of poetic and artistic conception of poetic space. Our ancestors of the traditional thinking habits are mostly heavy intuition, experience, to integrity, comprehensive mode of thinking. In the whole, comprehensive thinking mode, which specializes in classical poetry creation, is characterized by simple would be complicated language, dapper style and rich implication between the height of the conjunction. All works of artistic conception, should both do scene and do the boundary begets like outside, false or true which give a person leave a imagination that several layers of meaning are derived from reality to develop aesthetic imagination. Artistic conception is the land born like begets the outside, the actual product. Generally speaking, the virtual environment is the sublimation of reality, it reflects is habitat creation intention and purpose, embodies the artistic conception of art taste and aesthetic effect, restricting is creation and description, was the leader of the soul and structure of artistic conception. But, small virtual environment can be born out of thin air, in the artistic conception creation, everything must also implement to the reality of the figures.

**The Clustering based Classification Algorithms for Conception.** Most of existing text feature selection methods are serial and are inefficient timely to be applied to Chinese massive text data set. So, it is a hotspot of text mining how to improve efficiency of text feature selection by means of parallel thinking. Combining genetic algorithm with parallel collaborative evolutionary, a parallel collaborative evolutionary genetic algorithm based on rough sets was designed and used to select text features. Each feature there is a certain tendency, namely when reading a poem poetry works, through to the individual characteristics of artistic conception of experience and readers will be easier to work with certain category associated to express emotion. Is this effect, it is because if the characteristics of a work item one aspect of the artistic conception, environment readers will naturally associated with this aspect. The type string similarity calculation method is shown below.

$$\cos(W_1, W_2) = \frac{\sum_{i=1}^n w_{1i} \times w_{2i}}{\sqrt{\left(\sum_{i=1}^n (w_{1i})^2\right) \times \left(\sum_{i=1}^n (w_{2i})^2\right)}} \quad (5)$$

Hierarchical clustering method is also a kind of common clustering algorithm and for the need for clustering of data according to the level of the steps. According to the different methods of

hierarchical decomposition, hierarchical clustering method can be divided into a bottom-up and top-down clustering. For every feature of the artistic conception with the characteristics of a certain tendency of feature representation method is improved, puts forward the mode with vector to represent the feature, vector of each component represents the feature weight in each category. After the analysis from the perspective of the reader, the reader can feel the song poetry as a whole by the artistic conception, and from the conception to convey the author wants to convey thoughts and feelings, based on this proposed to the characteristics of clustering, on the one hand, those with similar artistic feature together, to achieve the purpose of semantic content retain more poetry, on the other hand, the clustering vector can be dimension reduction, improve the classification performance of the algorithm. Mean shift is calculated at the current point as the center and the most populous area around it and then move the center of the current the most heavily populated areas after repeated iterative calculation when the center was no longer drift or drift range within the scope of the threshold value and then you can get the largest local density of points.

### Experiment and Simulation Result

In this section, we conduct numerical simulation on the proposed methodology. The data is captured from the google scholar search engine. We use the standard library to test the effectiveness of the proposed approach. The experimental result indicates that our method performs well.

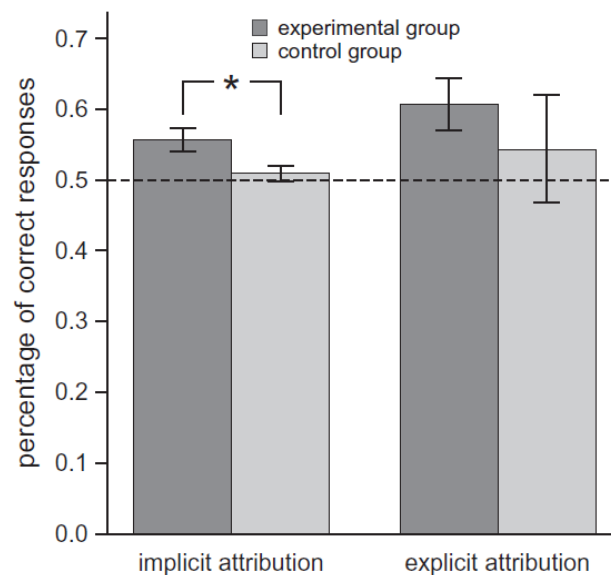


Fig. 1. The Experimental Result for the Proposed Algorithm

### Conclusion and Summary

In this paper, we conduct theoretical analysis and experimental implementation of the classification algorithms for the classical poetry artistic conception based on feature clustering methodology. Through sort of classical poetry into the problem of text categorization we can make use of some classical machine learning method, compared with general text but Chinese classical poetry has a special characteristic especially compared with the modern text difference is bigger. Based on statistical classification method is, in essence, a quantitative reasoning under uncertainty. It through the statistics and analysis on the large-scale corpus information needed to gain knowledge so that they can on the whole document of after data processing provides more objective data quality with high coverage and good consistency. The result proves the effectiveness of the proposed methodology and in the future, we plan to conduct more research to polish the current algorithm.

## References

- [1] He, P., Xie, G., Salamatian, K., & Mathy, L. (2014). Meta-algorithms for Software-Based Packet Classification. IEEE International Conference on Network Protocols (pp.308 - 319). IEEE.
- [2] Trappey A J C, Trappey C V, Hsu F C, et al. A fuzzy ontological knowledge document clustering methodology[J]. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 2009, 39(3): 806-814.
- [3] Song Q, Ni J, Wang G. A fast clustering-based feature subset selection algorithm for high-dimensional data[J]. Knowledge and Data Engineering, IEEE Transactions on, 2013, 25(1): 1-14.
- [4] Boş R I, Heinrich A, Wanka G. Employing different loss functions for the classification of images via supervised learning[J]. Central European Journal of Mathematics, 2014, 12(2):381-394.
- [5] Keylock C J, Singh A, Venditti J G, et al. Robust classification for the joint velocity-intermittency structure of turbulent flow over fixed and mobile bedforms[J]. Earth Surface Processes & Landforms, 2014, 39(13):1717–1728.
- [6] Boutemedjet S, Bouguila N, Ziou D. A hybrid feature extraction selection approach for high-dimensional non-Gaussian data clustering[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2009, 31(8): 1429-1443.
- [7] Evenbly G, Vidal G. Algorithms for entanglement renormalization: boundaries, impurities and interfaces[J]. Journal of Statistical Physics, 2013, 157(4-5):931-978.