International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.10, (2016), pp.295-304
http://dx.doi.org/10.14257/ijsip.2016.9.10.28

Research on Traditional Chinese Painting Classification Algorithm Based on Support Vector Machine

Ximan Shi

Art college of Xinxiang University, Xinxiang, Henan 453002, China
shiximan@sina.com

Abstract

Along with the advancement of Internet and digital technologies, more and more Chinese traditional paintings are becoming available on the Internet. As a result, computerized indexing and classification of given Chinese paintings emerge to be one of the focused research areas over recent years. As traditional Chinese paintings rely on the special drawing tools to illustrate the artistic styles, it distinguishes from western paintings in terms of strokes, contours, color and textures etc. Additionally, drawing lines play important roles in most traditional Chinese paintings. Consequently, the existing research on Western paintings is normally not applicable to traditional Chinese paintings. In addition, color-based approaches are also not applicable as traditional Chinese paintings mostly rely on gray scale texture to express their art styles and content. This thesis reports my intensive research program on computerized classification and automated learning and analyzing techniques for traditional Chinese paintings, in which a number of novel research and ideas are developed and put forward for style-based classification as well as its related theories and new concept introductions. My own novel contribution can be highlighted in machine learning, especially one-class SVM (OCSVM) based classification of traditional Chinese paintings. In this paper, the one-class SVM technology is revised to introduce a supervised learning element and arranged into a parallel OCSVM classifier. Based on the statistics features, a new concept of enforced learning has been introduced to remove the false positives at each learning cycle together with a new scheme of adaptive upgrading of decision parameters. Extensive experimental results support that the proposed new classifier achieves significant improvements in comparison with existing representative techniques.

Keywords: *Art Classification; Traditional Chinese Painting (Ink-wash Paintings) Analysis; Wavelet Transform; Feature Extraction; One-class SVM*

1. Introduction

Chinese painting [1-2], as the essence of Chinese culture and art, has begun to move towards the international market. More and more museums, art galleries, cultural relics shops, and even individuals began to choose the Chinese painting as a collection or investment objects. But the traditional painting takes rice paper and silk as material so that difficulty in the management, collection and maintenance and the cost is very high, which will undoubtedly hinder the pace of internationalization development of Chinese painting [3-4]. With development of network and digital, more and more of the calligraphy and paintings appeared in the network, especially the museum collections of the construction of digital archives. Additionally, many organizations also collect digital image data of the art collection for people's reference on the Internet. Anyone can access a wide variety of art treasures through the Internet, and more and more artists try to show and sell their products on the Internet. Let traditional Chinese painting be digitized, and proceed computer storage and management, which will bring great convenience to collectors and institutions [5]. They can put Chinese painting on the Internet to display, manage, and

ISSN: 2005-4254 IJSIP

auction. Therefore, the digitization of Chinese painting has become a trend. How to facilitate the retrieval and classification of specific images has become a hot research topic. In particular, how to make use of the computer to analyze the different artist works from such a large number of art database, and compare their style, become a difficult issue. China painting is different from western art, whose formation relies on Chinese brush, ink, water and color. Ink color shade variation can be expressed by the gray level changes of pen, so Chinese painting usually relies on strokes to make outline, texture, volume to represent the style of an artist. While western paintings emphasize on utilizing light to present the activity performance of object, which is one of the important means of modeling of Western painting [6]. As the Chinese painting is mainly based on the line, it makes the Chinese and Western painting are of obvious variety in the modeling means. In consequence, some of the existing researches on the western painting cannot be directly used in the analysis of ink painting. In addition, because the ink painting is not colored or less colored and most of them have no rich color performance, some of existing extraction and classification algorithm based on the feature is not suitable for the study of ink painting [7]. In summary, the existing literature and research results cannot be directly used for the analysis and classification of the artistic style of Chinese painting. It has great significance if the computer classification and retrieval technology are utilized, extract features of artistic style of painting through image processing means so as to intelligently recognize and classify the painting style of famous ink.

2. Further Research on Automatic Classification of Chinese Ink and Wash Painting by Support Vector Machine

In this chapter, we focus on the research on commonly used support vector machines and related theories and algorithms in exploration of machine learning, and apply it to the computer automatic classification of ink painting, so as to find a more effective classification algorithm for machine learning. It lays a foundation for further research on the direction of machine learning. Starting from the basic support vector machine theory and concept, this chapter introduces the research work to further explore the current international advanced algorithm about support vector machine (SVM), introduces a so-called class support vector machine (one class SVM, OC-SVM). And use its robustness against the noise to design a parallel class support vector machine algorithm to complete the automatic classification of the ink and wash painting. At the same time, the paper further proposes an interactive algorithm of using statistical features to reduce the false positive rate, to further improve the classification results of parallel one class support vector machine. Therefore, the innovation points of this chapter can be summarized as two points. The first one is that a new kind of parallel support vector machine ink painting classification algorithm is proposed; the second is that an interactive algorithm based on statistical features to reduce the false positive rate is proposed.

2.1. Main Concepts of Support Vector Machines

Support vector machine is usually divided into two kinds of linear and nonlinear. The former, as an entry concept for support vector machines, provide a good basis for the understanding of the concepts, its basic concept equals to a given N training vector $\{x_1, x_2, \ldots x_N\}$. The corresponding output of each vector can be indicated by $y_i=\{y_1, y_2, \ldots y_N\}$. While we need to find a hyper-plane $g(x)=w^T x+w_0$ to get the following classification result:

$$x_i \in w_I : w^T x + w_0 \geq 1 \quad (y_i = 1) \tag{1}$$

$$x_i \in w_I : w^T x + w_0 \leq -1 \quad (y_i = -1) \tag{2}$$

And w= (w_1, w_2,…w_N) is taken as support vector. Therefore the next step is to determine the parameters w and w_0 from the given N training vectors. Using the main concepts of linear support vector machines, we can get the optimal hyper plane (line) expressed as x_1=0. The reason is that y_1=y_2=1; y_3=y_4=-1. The corresponding support vector machine can be displayed in Figure 1.

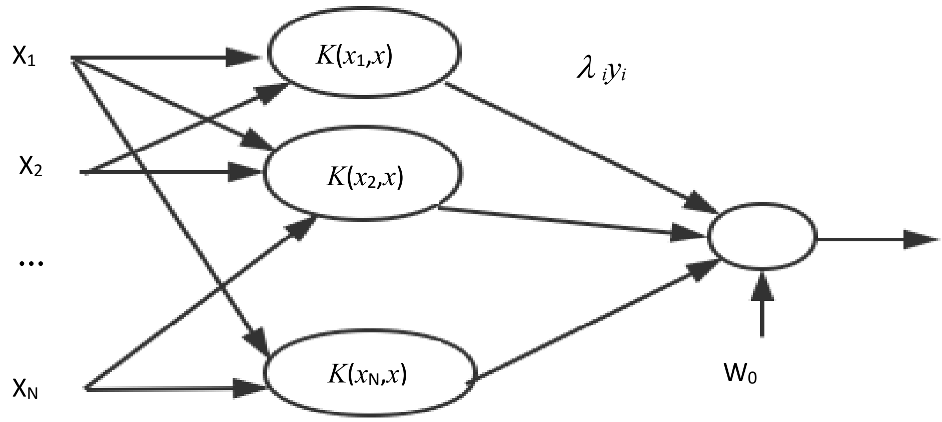

Figure 1. Route Chart of Non-linear Support Vector Machine

2.2. Classification Method of Parallel Support Vector Machine to Ink Painting

Although numerous studies have shown that support vector machines can achieve very good classification results, its main weakness is the need of pre-labeled supervised learning, and it is very sensitive to the noise. In other words, a small fraction of the labeled data is likely to have a serious negative impact on its classification performance. As a result, in 1999, Scholkopf and others scholars proposed the extension of the general support vector machine into a class of support vector machines (SVM one-class, OC-SVM). Their main work can be summarized as following two points.

The first is to change the supervised learning adapt to the labeled training data in the original support vector machine to no-tag, that is, unsupervised learning. The second is, when reflecting the input data to a feature space, adjust the corresponding kernel function so that the majority of the training data are mapped to a maximum possible envelope space, and a few are concentrated in the vicinity of the origin. Then a binary classification problem can be solved by the method of a class of dominant, as shown in Figure 2. This method is well applied in the computer network security, especially the characteristics of the uncertainty and failure in obtaining the label in advance in network security attack data. The one class support vector machine classification method can well detect suspicious events, also anomaly, and produce timely warning for a computer system.

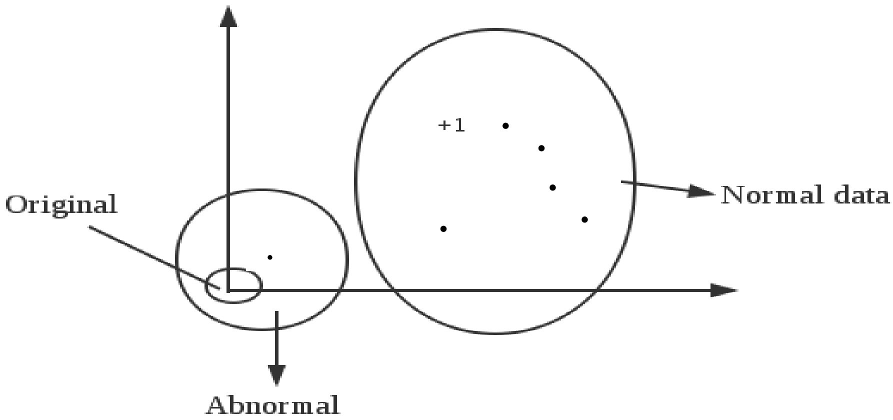

Figure 2. Fundamental of One-class Support Vector Machine

However, there are still some problems needing to be solved to make use of the advantages of this kind of support vector machine in the anti noise to do the classification of ink painting. They specifically include: (1) training data becomes targeted weak because of no prior label, which leads to the classification accuracy is not high; (2) the method of choosing the corresponding decision parameters from the default value can not reflect the flexibility of the style of Chinese painting and thus a low adaptability. To solve these problems, starting from the main concept of a self-sustaining vector machine, compare a painter's painting to other artists' works, using asymmetric relationship among the numbers to map other painter works to a class of space in training. While the work of a specific artist is seen as outlier mapped to the small space taking the origin as the center. Taking this kind of support vector machine as the basic unit, construct a multi machine parallel classifier, as shown in Figure 3.

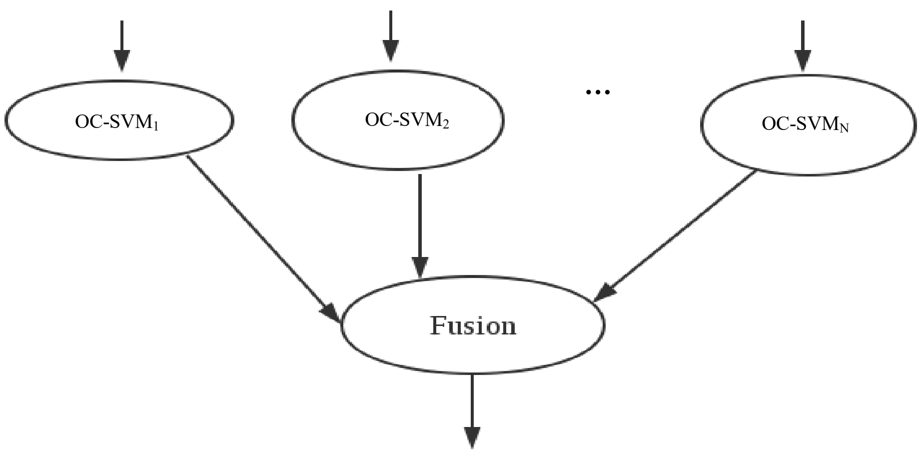

Figure 3. Structure Diagram of Parallel One-class Support Vector Machine

The corresponding parallel adaptive support vector machine classifier's interactive training process is shown in Figure 4.

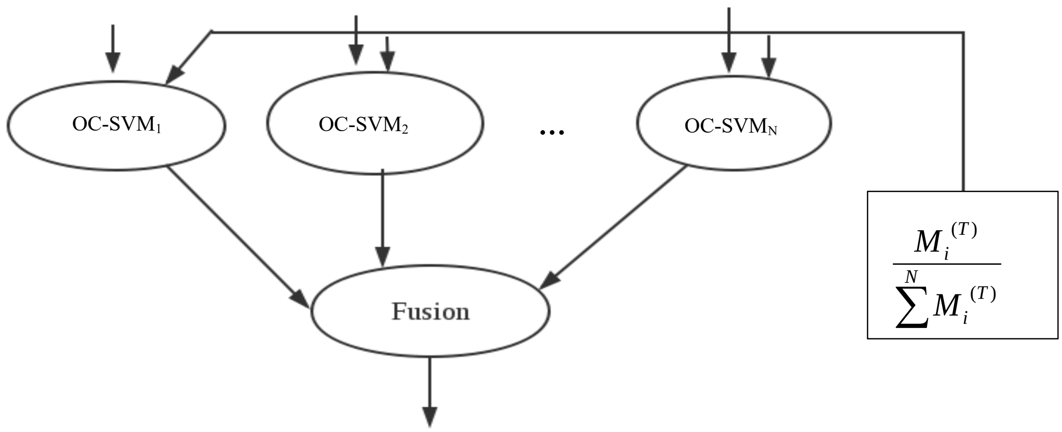

Figure 4. Process and Structure of Parallel One-class Support Vector Machine Classifier

2.3. Interactive Algorithm for Reducing the Rate of False Positive Based on Statistical Features

In order to further improve the classification performance of parallel support vector machines shown in Figure 24, an enhanced SVM learning training algorithm based on statistical features is proposed to reduce the false rate. The specific method is for the given training data, divide the results into two sets according to their original label. $\psi_{TP} = \{x_i^{TP} | i \in [1, N_{TP}]\}$ and $\psi_{FP} = \{x_i^{FP} | i \in [1, N_{FP}]\}$, in which TP is shorted by True Positive, means the right classification result; FP is shorted by False Positive, means the false classification result. x_i^{TP} and x_i^{FP} is correspondingly the input vectors of the right classification result and the false classification result respectively. According to the specific learning of support vector machine, calculate the average value $\mu(v_i^{TP})$ and $\mu(v_i^{FP})$ of the specific value of the corresponding subset ψ_{TP} and ψ_{FP} of parameter v. Then determine a threshold to re-combine the classification results, the specific threshold determination is calculated by the expression in the following.

$$\begin{cases} \dfrac{\max}{i \in [1, N_{TP}]} (v_i^{(TP)}) \quad if \quad \mu(v_i^{(TP)}) \le \mu(v_i^{(FP)}) \\ \hline \dfrac{\min}{i \in [1, N_{FP}]} (v_i^{(FP)}) \quad else \end{cases} \tag{3}$$

N_{TP} and N_{FP} represents the number of paintings judged by the right classification result and the false classification result respectively. If $\mu(v_i^{TP}) \leqq \mu(v_i^{FP})$, it indicates a tendency that the corresponding parameter v of all the input vectors in the subset of ψ_{TP} is smaller than that in ψ_{FP}. Therefore, when the threshold T=max(v^{TP}), it is possible to correct the classification faults smaller than T. In other words, in the subset ψ_{FP}, only the false judgment of $v^{FP} > T$ is concluded. As a result, the false judgment of $v^{FP} < T$ can be reduced and has less influence on the original right classification result. The process of reducing the false rate is shown in Figure 5. In the figure, the majority of TP points are in the beneath. Consequently, for the points in ψ_{FP}, only the corresponding point FP on the condition of $v>T$ is considered, while the FP points smaller than T can be corrected to ψ_{TP}.

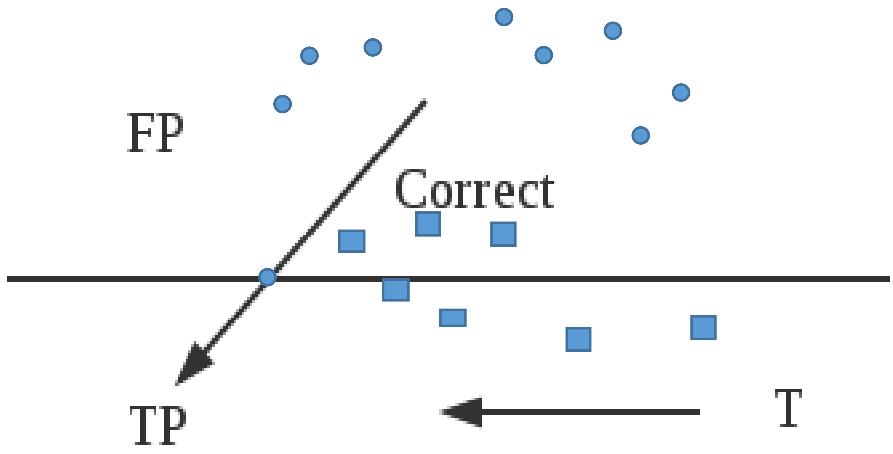

Figure 5. Correction of Erroneous Judgment

In a similar way, if $\mu(v_i^{TP}) > \mu(v_i^{FP})$, it means that the corresponding parameter v of all the input vectors in the subset of ψ_{TP} is larger than that in ψ_{FP}. In consequence, if the threshold T is the smallest value v in ψ_{TP}, it is possibly that the FP points larger than T belong to ψ_{TP}. Eventually, if the corresponding value v of FP points are larger than T, then they can be corrected and belonged to ψ_{TP}.

2.4. Experiment Result and Analysis

To objectively assess and verify the feasibility of the new concept introduced in this chapter and the classification algorithm of the parallel one-class support vector machine, we adopt the images in a unified set of ink painting in the database as initial training and test data to test and experiment the classification algorithm and parallel one-class support vector machine classifier. In order to keep the consistency of the whole paper, we use the same database. Among them, all the ink and wash paintings are divided into two equal parts. One part is taken as the training data to set the parameters needed for the SVM classifier. The other part is regarded as the test data so as to ensure that in the actual test, the input ink painting isn't learned by the classifier.

We also used three different cases to complete our experiments, namely case 1, case 2 and case 3, and the corresponding experimental results are summarized in Table 1. Except for the algorithms proposed by ourselves, two algorithms did and published by other people are concluded. That is method of MHMM and C4.5 decision tree. In consequence, we can do a comparative test in allusion to the algorithm we proposed. And then we get the advantages and disadvantages of some new concepts and algorithms proposed in this chapter compared with the current research results of the same kind.

Table 1. Classification Result of Chinese Ink Painting

Case 1						
Precision ratio p **Recall ratio** γ	**OC-SVM**		**MHMM**		**C4.5**	
	p(%)	γ *(%)*	*p(%)*	γ *(%)*	*p(%)*	γ *(%)*
Huang Gongwang	95.2	100	90	90	88.9	80
Zheng Banqiao	100	95	90	90	81.8	90
Average value	**97.6**	**97.5**	90	90	85.4	85
Case 2						
Precision ratio p **Recall ratio** γ	**OC-SVM**		**MHMM**		**C4.5**	
	p(%)	γ *(%)*	*p(%)*	γ *(%)*	*p(%)*	γ *(%)*
Xu Beihong	95	95	85	85	76.5	65
Wu Changshuo	90.5	95	77.3	85	72.7	80
Huang Gongwang	95.2	100	83.3	75	71.4	75
Zhen Banqiao	100	90	80	80	80	80
Average value	**95.2**	**95**	81.4	81.3	75.2	75
Case 3						
Precision ratio p **Recall ratio** γ	**OC-SVM**		**MHMM**		**C4.5**	
	p(%)	γ *(%)*	*p(%)*	γ *(%)*	*p(%)*	γ *(%)*
Xu Beihong	100	95	83.3	75	60.9	70
Wu Changshuo	90	90	77.3	70	70.6	60
Huang Gongwang	95.2	100	89.5	85	58.3	70
Zhen Banqiao	100	90	72.7	80	68.4	65
Liu Danzhai	86	95	61	70	65	55
Average value	**94.3**	**94**	76.8	76	65	64

From the experimental results of Table 1, our algorithm has a great advantage compared with the MHMM and C4.5 decision tree method. In all the three experiments, our algorithm precision and recall ratio are higher than the two algorithms compared, from which we can see that our scientific work is in the leading position. In addition, the results of Table 1 experiments also show that with the increase of the number of painters, the classification results will gradually become worse. But compared with the other two methods, the extent of being worse is much smaller. We can see that the robustness of the algorithm is better than that of the existing classification schemes. To further confirm the robustness of our algorithm, in the classification experiment, we add an unknown painter. And some paintings that have been confirmed not belonging to the 5 painters in the database are used to complete the work of the SVM classifier training. The experimental results obtained are presented in Table 2. The experimental results show that the classification results related to Zheng Banqiao have an obvious reduction. This is because the majority of the increased ink painting of the "unknown painter" category is close to that of Zheng Banqiao. We deliberately make this arrangement in the experiment so that the experimental conditions are more harsh to confirm the robustness of the algorithm tested.

Table 2. Experiment Result of Testing Robustness

Artists	**Precision ratio** p	**Recall ratio** γ
Xu Beihong	95.2	100
Wu Changshuo	90	90
Huang Gongwang	95	95
Zhen Banqiao	80	80
Liu Danzhai	95	95
Unknown artists	72.2	65
Average value	87.9	87.5

The experimental results discussed above show that the proposed parallel class support vector machine classifier has a good grasp of the different art styles used by the artists. But for the ink and wash painting in similar style, there is still much room for further research on making accurate classification.

Finally, we also propose an enforced SVM training learning method. It makes use of the classification results of the training data for further statistical analysis and correct the false positives on or under the threshold to improve the classification results. In order to verify the interactive enforced training scheme and the relevant new concept, we have made a further study on the case 3, and the corresponding experimental results are shown in Table 3. Thus, the interactive reinforcement learning and training program we proposed have a certain improvement on our previous parallel support vector machines, which confirms the effectiveness of our research work.

Table 3. Comparison of Experiment Result of Interactive Enforced Training Algorithm

Precision ratio p Recall ratio γ	OC-SVM		Interactive Enforced Algorithm		MHMM		C4.5	
	$p(\%)$	$\gamma(\%)$	$p(\%)$	$\gamma(\%)$	$p(\%)$	$\gamma(\%)$	$p(\%)$	$\gamma(\%)$
Xu Beihong	100	95	100	100	83.3	75	60.9	70
Wu Changshuo	90	90	95.2	100	77.8	70	70.6	60
Huang Gongwang	95.2	100	100	100	89.5	85	58.3	70
Zhen Banqiao	100	90	100	95	72.7	80	68.4	65
Liu Danzhai	86.4	95	95	95	60.9	70	64.7	55
Average value	94.3	94	98.04	98	76.8	76	64.6	64

3. Conclusions

As an important cultural heritage and Chinese art treasure, Chinese paintings have aroused a great deal of attention in the contemporary art world, and it is even used as a treasure in the collection field. Because the paper picture is not easy to save, painting digitization is an irresistible trend [8]. More and more works of painting and calligraphy appear in the network. Museum begins to construct the digital archives, and many organizations collect image data for people's reference on the Internet. Their computerized management, including indexing, retrieval, analysis and classification is becoming an important research topic in many research fields [9-10].

Chinese painting is different from western painting, so some of the existing researches on western painting can not be directly used in the analysis of Chinese ink painting. Therefore, based on the automatic analysis of the art style of ink painting, this paper studies the automatic classification of Chinese ink painting. This paper presents a theoretical and technical framework for the classification of the art style of Chinese ink painting. The main idea can be summarized by a simple image processing flow, as it is shown in Figure 6.

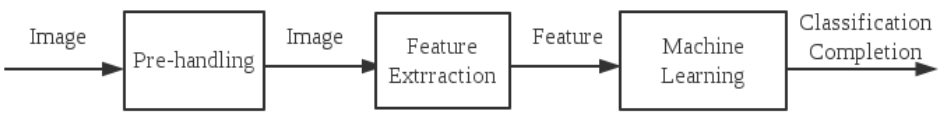

Figure 6. Flow Chart of Image Handling

In addition, a classification method based on parallel support vector machine is proposed in this paper. The proposed method is based on the interactive training algorithm of statistical feature reduction and the adaptive parameter decision algorithm, which gets a good classification result and has robustness. Finally, in order to take into account the overall image and local art style, this paper proposes a fusion algorithm. In the premise of maximizing the art style, fusion algorithm combines with the overall and partial class results advantage, eventually achieves the automatic classification of ink painting based on the art style, and outputs the classification results.

Acknowledgments

The author is indebted to the referees for nice suggestions.

References

[1] T. Simon and D. Koller, "Support vector machine active learning with applications to text classification", Journal of machine learning research, vol. 2, **(2011)**, pp. 45-66.
[2] P. H. Reza, "Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province", Iran. Journal of Earth System Science, vol. 122, no. 2, **(2013)**.
[3] S. F. Terrence, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data", Bioinformatics, vol. 16, no. 10, **(2012)**.
[4] M. W. King and P. A. Resick, "Data mining in psychological treatment research: A primer on classification and regression trees", Journal of consulting and clinical psychology, vol. 82, no. 5, **(2014)**, pp. 895
[5] C. Olivier, "Training a support vector machine in the primal", Neural computation, vol. 19, no. 5, **(2012)**.
[6] N. Singhal and M. Ashraf, "Performance enhancement of classification scheme in data mining using hybrid algorithm", Computing, Communication & Automation (ICCCA), 2015 International Conference on. IEEE, **(2015)**, pp. 138-141.
[7] N. Abtin and S. Hartmann, "Fault Classification of a Centrifugal Pump in Normal and Noisy Environment with Artificial Neural Network and Support Vector Machine Enhanced by a Genetic Algorithm", International Conference on Theory and Practice of Natural Computing. Springer International Publishing, **(2015)**.
[8] T. Cheng, "The Interpretation of Huang Binghong's Accumulated Ink Painting Style", Art and Design, vol. 10, **(2013)**.
[9] W. Hung, "Transcending the east/west dichotomy: A short history of contemporary Chinese ink painting", New York: Metropolitan Museum of Art, **(2013)**.
[10] Q. Lv and H. Hu, "Research on parallel computing model and classification algorithm based on data mining process", International Journal of Security and Its Applications, vol. 9, no. 5, **(2016)**, pp. 231-240.

Authors

Ximan Shi, she was born in 1982. She received the Bachelor's degree in school of art from Henan University, in 2004, and received the Master degree in school of art from Xiamen University, in 2009. Currently, she is a lecturer at art college of Xinxiang University. Her research interests include Internet of artistic designing and Classification algorithm.