2019 IEEE 4th International Conference on Signal and Image Processing

Overview of one-Class Classification

Sun Wenzhu, Hu Wenting, Xue Zufeng, Cao Jianping
Naval Aeronautical University Qingdao Branch
Qingdao, China
e-mail: sunwenzhulm@163.com, 670939743@qq.com, 1643608947@qq.com, jp_cao2016@126.com

***Abstract*—One-class classification, which was tested successfully in unbalanced sample classification problems, is one of the hotspots in pattern recognition research. This paper first analysis the commonly used one-class classification methods, then classifies these methods into three categories: boundary based method, re-construction based method and border based method. Finally, a series of testing experiments based on artificial database are designed to test the advantages and disadvantages of these methods from the aspect of learning ability, classification decision and algorithm complexity.**

Keywords-one-class classification; pattern recognition; unbalanced sample

I. INTRODUCTION

Unbalanced sample classification problems exists widely in our lives. For example, in the problems of mechanical fault diagnosis [1], network intrusion detection [2], medical diagnosis [3], [4], most of the collected samples are normal samples, and abnormal data rarely occurs. In the problems of face detection, target retrieval [5] and character detection [6], although the abnormal samples are easy to obtain, but there are too many types of abnormal samples, and it is almost impossible to obtain all of them. For example, in face detection, all non-human faces can be used as abnormal samples, and all individual samples are not representative. In the pattern recognition process, the traditional solutions mainly include oversampling [7], [8], downsampling [9], [10], and sample weighting [11], [12]. However, when the problem of unbalanced sample number is serious, these classification methods cannot obtain high classification accuracy. Based on this, Moya proposed the idea of one-class classification [13] in 1996. In 2001, Tax summarized and elaborated the one-class classification in his doctoral thesis and proposed support vector data description (SVDD) [14], [15], which indicates that one-class classification has become an important branch of pattern recognition.

The classification idea of the one-class classification method is different from the traditional two-class classification (multi-class classification problems can be equivalent to the combination of multiple two-class classification problems, so the two-class classification is the basis of the classification problem), in the traditional two-class classification, all samples are given a category label of "+1" or "-1".

The training process of the classifier is the process of forming the classification surface between the two-class classification samples, as shown in Figure 1. Figure 1(a) shows the results of two-class classification under the condition of balanced samples. At this time, the classification surface is located between the two-class classification samples. Figure 1(b) shows the results of two-class classification under the condition of unbalanced samples . When the two-class classification sample size difference is large, the classification surface of the traditional classification method (support vector machine in the experiment) is obviously shifted to the less sample side, resulting in lower classification accuracy.

In the one-class classification problem, there are many samples with a “+1” category label (target samples) and few samples with a “-1” category label (non-target samples). The one-class classification can form a boundary description of the target samples only by using target samples during the training process, as shown in Figure 1(c). The boundary description here is similar to the concept of "envelope", that is, using a closed set containing the distribution area of most training samples. During the test, the samples within the boundary description are classified as target samples, and the samples outside the boundary description are classified as non-target samples. In this way, all samples that differ greatly from the target samples are classified into non-target samples, which improves the classification accuracy.

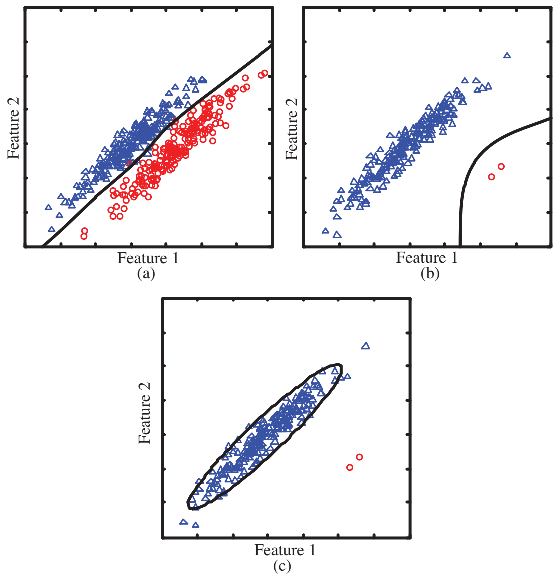

Figure 1. Results of two-class classification

II. METHOD INTRODUCTION

Since the one-class classification method is based on the assumption that only the target samples are available, but

978-1-7281-3660-8/19/$31.00 ©2019 IEEE

because of the lack of non-target samples, it cannot determine whether the boundary description of the training is over-learning or under-learning [16]. Therefore, in the evaluation of the results of the one-class classification method, it is usually assumed that the implicit non-target samples are evenly distributed, so the one-class classification problem is usually transformed into the target sample domain description density level estimation problem [17] or the target sample boundary description minimization estimation problem[18]. At present, the one-class classification methods are mainly divided into three types: density-based methods, reconstruction-based methods, and boundary-based methods.

(1) Density-based one-class classification method

The density-based one-class classification method first estimates the target set sample distribution density, and then sets a threshold, and the region where the sample distribution density is higher than the threshold is used as the acceptance domain of the test sample. Commonly used one-class classification methods are: Gaussian model, mixed Gaussian model [19] and Parzen density model.

(2)Reconstruction-based one-class classification method

The reconstruction-based one-class classification method establishes a sample generation model by using training samples as prior knowledge. Test samples can be described as a state of the generated model. Commonly used reconstruction-based one-class classification methods are: k-center, k-means, PCA one-class classification [20] and self-organizing mapping method [21].

(3) Boundary-based one-class classification method

The boundary-based one-class classification method directly optimizes the boundary description of the target samples based on prior knowledge. Commonly used boundary-based one-class classification methods are: k-nearest neighbor method, one-class support vector machine, (OCSVM) and SVDD.

III. Comparative Analysis of One-class Classification Methods

In order to compare the performance of each one-class classification method, we discuss the performance indicators of the methods separately mentioned in Section 2. Since SVDD and OCSVM are theoretically uniform and SVDD is more general [15], only the experimental results of SVDD are retained in the experiment.

A. *Learning Ability and Generalization Ability*

Learning ability and generalization ability can be seen from the visual boundary description of the classification method. We apply the 9 methods to the 2D Gauss data set and Banana data set generated by the *prtools* toolbox (http://www.prtools.org). On the set, the distribution parameters of the Gauss data set are:

$$\boldsymbol{\mu} = [0, 0], \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix} \tag{1}$$

where: $\boldsymbol{\mu}$ is the mean, $\boldsymbol{\Sigma}$ is the covariance matrix, the resulting boundary description is shown in Figure 2 and Fig. 3, respectively, the dots are the training samples, the solid line is the boundary description of the method.

It can be seen from Fig. 2 and Fig. 3 that the boundary description of the GM method is a super ellipsoid, which is suitable for Gauss dataset classification, but the classification effect is not ideal on the Banana dataset. This shows that the GM method is only applicable to the one-class classification with the convex distribution samples. The other two density-based one-class classification methods, MOG and Parzen window have more flexible boundary descriptions, and have better classification effects on both data sets. The figure shows that the boundary description of the MOG method is smoother than the boundary description of the Parzen window. This is because the MOG method first unsupervised clustering all samples into clusters, each cluster supporting a Gaussian density window. When $k = N$ (N is the number of training samples), the MOG method is equivalent to the Parzen window method; When $k = 1$, the MOG method is equivalent to the GM method. When the value of k is appropriate, the MOG method can obtain faster training speed, higher classification accuracy and better generalization ability. However, as a density-based one-class classification method, MOG will be subject to “Dimension disaster” in small sample high-dimensional classification. and cannot establish accurate density estimates. The classification results of the k-center and k-means methods are similar, but the boundary description of k-means is smaller. This is due to the difference in the error functions used by the two methods, the k-means takes the "majority" samples into account and the k-center takes the "worst" samples into account. The boundary description of the PCA method is linearly open, with strong generalization ability but poor learning ability. The large boundary description of the SOM and k-nearest neighbor methods indicates that the two methods can achieve less expected risk, but at the same time have over-learning problems. The SVDD method accurately describes the sample distribution of the two data sets and the boundary description is compact and smooth, indicating that the SVDD method has better learning ability and generalization ability.

B. *Algorithm Classification Accuracy*

In order to compare the classification accuracy of each one-class classification method, we give the results of the area under the receiver operating curve (AUC) of the 9 methods in the 2D Gauss dataset and the Banana dataset [22]. ,As shown in Table I. In the experiment, the data set generation function generates a class of 2000 samples as target samples, of which 1000 target samples are used for training, and the remaining 1000 target samples are used for testing. The non-target samples used for testing are the 1000 samples evenly distributed in the sample space [23]. It can be seen from Table I that the classification accuracy of the MOG method, Parzen window, k-means and SVDD methods is higher on both data sets, while the GM method performs better on the Gauss data set than the Banana data set. this is because the GM method can only form a Gaussian form of boundary description, and the Gaussian form of the boundary description is suitable for the Gauss data set. The k-center, PCA, SOM, and k-nearest neighbor methods have lower AUC values due to defects in learning ability or generalization ability.

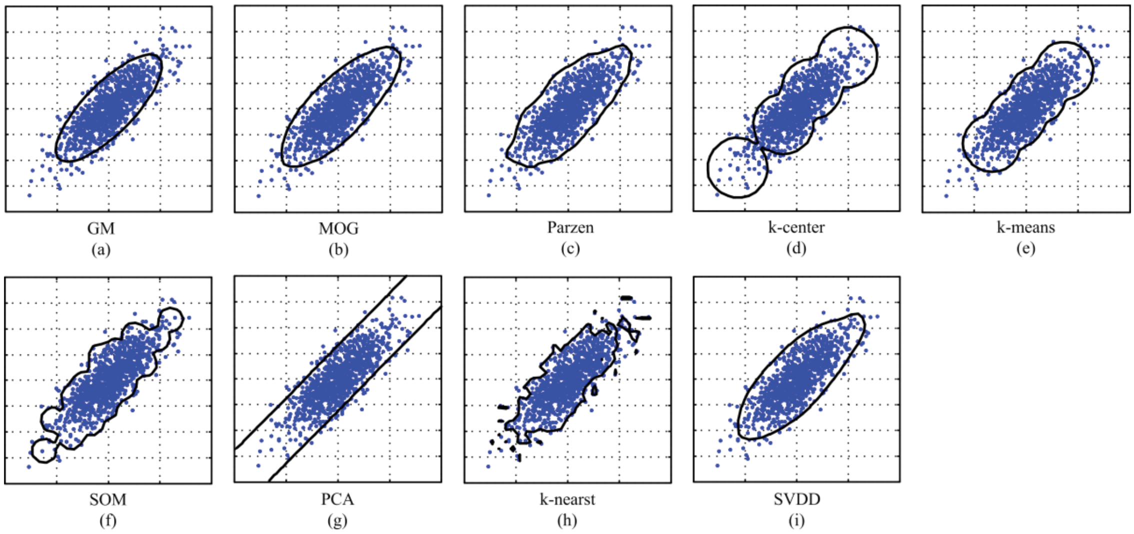

Figure 2. The results of one-class classification on the Gauss data set.

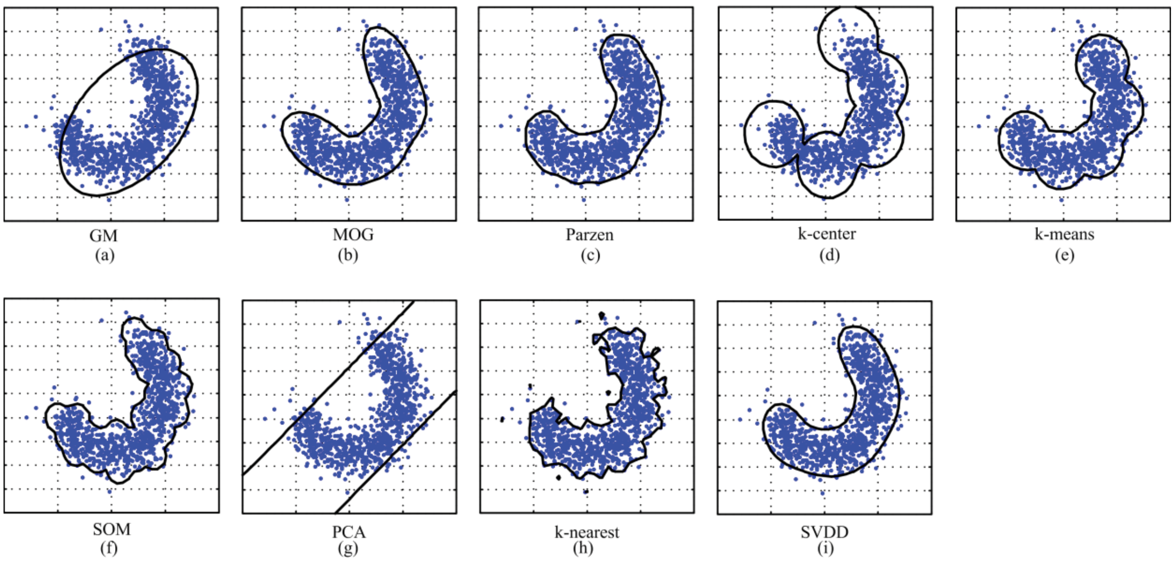

Figure 3. The results of one-class classification on the Banana dataset

TABLE I. COMPARISON OF AUC VALUES FOR EACH METHOD ON GAUSS AND BANANA DATA SETS

Methods	**Gauss Data set**	**Banana Data set**
GM	0.9437	0.8588
MOG	0.9434	0.9165
Parzen Window	0.9423	0.9165
k-center	0.8872	0.8696
k-means	0.9313	0.9101
PCA	0.8855	0.7682
SOM	0.9138	0.8911
k-nearest	0.9173	0.8923
SVDD	0.9434	0.9168

TABLE II. ALGORITHM COMPLEXITY OF ONE-CLASS CLASSIFICATION METHOD

Methods	Training time complexity	Training space complexity	Testing time complexity	Testing space complexity
GM	$O(ND^2)$	$O(D^2)$	$O(1)$	$O(D^2)$
MOG	$O(ND^2)$	$O(D^2)$	$O(1)$	$O(D^2)$
Parzen Window	$O(ND^2)$	$O(N^2D^2)$	$O(N^2D^2)$	$O(ND^2)$
k-center	$O(N)$	$O(N)$	$O(1)$	$O(1)$
k-means	$O(N)$	$O(N)$	$O(1)$	$O(1)$
PCA	$O(N)$	$O(D^2)$	$O(1)$	$O(D^2)$
SOM	$O(Nk_{knots})$	$O(Nk_{knots})$	$O(k_{knots})$	$O(k_{knots})$
k-nearest	N/A	N/A	$O(N^2)$	$O(N^2)$
SVDD	$O(N^3)$	$O(N^2)$	$O(N_{SV})$	$O(N_{SV}^2)$

Note: k_{knots} is the number of nodes in the SOM, N is the number of samples, D is the sample dimension.

C. *Algorithm Complexity*

Table II lists the algorithmic complexity of the 9 methods. In the GM method training process, the operation mainly focuses on solving the maximum likelihood parameter estimation with the number of samples, and establishing a covariance matrix in the memory, so the training time complexity is $O(N)$, and the training space complexity is $O(D^2)$. The MOG method is equivalent to the combination of the k-means method and k GM methods, and k generally takes a smaller natural number, so its complexity is identical in form to the GM method. All the parameters to be determined in the Parzen window method can be solved by the maximum likelihood method. Therefore, the training time complexity of the Parzen window method is less, but the test time complexity of the Parzen window method is large, because the boundary description needs all training samples, and during the test, Parzen needs to calculate the distance from the test samples to all training samples.

In the case of a large number of training samples, it will impose a large burden on the computer's arithmetic unit and memory. The k-means and k-center methods have the same complexity. In the training process, the convergence speed is fast, and less iterations can converge to the limit, so the time complexity is low. After the training is completed, only k samples are stored as the super-ball center and the threshold as the radius of the hyper-sphere, and so the required storage space is small. The PCA method mainly operates on the covariance matrix of the sample features, so the PCA method complexity is highly correlated with the sample dimension D. The training time complexity and spatial complexity of the SOM method are related to the number of nodes, and like other neural network methods, the error function of the iterative optimization process converges slowly, the number of iterations is more, and the training time is longer. The k-nearest neighbor method has no training process, but all training samples need to be stored. When the number of samples is large, a large storage capacity is required. The training time complexity of SVDD is the highest, because SVDD needs to solve the $O(N^3)$ complexity of the quadratic programming problem during the training process. However, the test time complexity and spatial complexity of SVDD are low, because the test time complexity and spatial complexity of SVDD depends only on the number of support vectors (N_{SV}), but N_{SV} is usually of the order of magnitude. Therefore, the SVDD method is generally suitable for offline training, classification of online tests.

IV. CONCLUSION

In this paper, the hotspot problem in pattern recognition in recent years, the one-class classification method, is deeply studied and given an overview. First of all, the background of the one-class classification method is analyzed, and the superiority of the one-class classification to the traditional two classification methods in solving the unbalanced samples classification problem is illustrated by experiments. Then, through the comparative analysis of various one-class classification methods, the existing one-class classification methods are divided into three types: density-based methods, reconstruction-based methods and boundary-based methods. Finally, through the experiments on the artificial dataset, the advantages and disadvantages of each method in terms of learning ability, classification accuracy and algorithm complexity are compared and the application range of each method is given.

REFERENCES

[1] Ashkan Moosavian, Hojat Ahmadi, Babak Sakhaei, and Reza Labbafi, Support vector machine and K-nearest neighbour for unbalanced fault detection[J]. Journal of Quality in Maintenance Engineering, 2014. 20(1): p. 65-75.

[2] Ciza Thomas, Improving intrusion detection for imbalanced network traffic[J]. Security and Communication Networks, 2013. 6(3): p. 309-324.

[3] Filippo Amato, Alberto López, Eladia María Peña-Méndez, Petr Vaňhara, Aleš Hampl, and Josef Havel, Artificial neural networks in medical diagnosis[J]. Journal of Applied Biomedicine, 2013. 11(2): p. 47-58.

[4] Wei-Liang Tay, Chee-Kong Chui, Sim-Heng Ong, and Alvin Choong-Meng Ng, Ensemble-based regression analysis of multimodal medical data for osteopenia diagnosis[J]. Expert Systems with Applications, 2013. 40(2): p. 811-819.

[5] Nenad Tomasev, Doni Pracner, Raluca Brehar, Milos Radovanovic, Dunja Mladenic, Mirjana Ivanovic, and Sergiu Nedevschi. Object recognition in wikimage data based on local invariant image features[C]. in Intelligent Computer Communication and Processing (ICCP), 2013 IEEE International Conference on. 2013: IEEE.

[6] Haixiong Fan, Fuxian Liu, and Lu Xia, Multi-label unbalanced classification algorithm based on probability LS-SVM[J]. Journal of PLA University of Science and Technology (Natural Science Edition), 2013. 14(2): p. 169-175.

[7] H. Han, W.Y. Wang, and B.H. Mao, Borderline-SMOTE:A new over-sampling method in imbalanced data sets learning, in Advances in Intelligent Computing. 2005, Springer: Berlin. p. 878-887.

[8] Hui Han, Wenyuan Wang, and Binghuan Mao, Over-sampling algorithm based on Adaboost in unbalanced data set[J]. Jisuanji Gongcheng/ Computer Engineering, 2007. 33(10): p. 207-209.

[9] Lin Shuyang, Li Cuihua, Jiang Yi, Lin Chen and Zou Quan, Under-sampling Method Research in Class-Imbalanced Data[J]. Journal of Computer Research and Development, 2011. 48: p. 47-53.

[10] Ashish Anand, Ganesan Pugalenthi, Gary B Fogel, and PN Suganthan, An approach for classification of highly imbalanced data using weighting and undersampling[J]. Amino acids, 2010. 39(5): p. 1385-1391.

[11] Zheng Enhui, Xu Hong, Li Ping and Song Zhihuan, Mining knowledge from unbalanced data based on v-support vector machine [J]. Journal o f Zhejiang University (Engineering Science), 2006. 40(10): p. 1682-1687.

[12] MAH Farquad and Indranil Bose, Preprocessing unbalanced data using support vector machine[J]. Decision Support Systems, 2012. 53(1): p. 226-233.

[13] M. Moya and D. Hush, network constraints and multi-objective optimization for one-class classification[J]. Neural Networks, 1996. 9(3): p. 463-474.

[14] Wenzhu SUN, Jianling QU, Yang CHEN, Yazhou DI, and Feng GAO, Heuristic sample reduction method for support vector data description[J]. Turkish Journal of Electrical Engineering & Computer Sciences, 2016. 24: p. 298-312.

[15] D. Tax, one class classification: Concept-learning in the absence of counter-examples. [D]. University of Delft, Netherlands

[16] P. Juszczak, Learning to recognise:A study on one-class classification and active learning[D]. Delft University of Technology, Delft

[17] R. Vert and J. Vert, Consistency and convergence rates of one-class SVM and related algorithms[J]. J. Mach. Learn.Res, 2006: p. 817-854.

[18] C. Scott and R. Nowak, Learning minimum volume sets[J]. Journal of Machine Learning Research, 2006. 7: p. 665-704.

[19] Christopher M. Bishop, Pattern Recognition and Machine Learning. 2006, Singapore: springer.

[20] S.L. Shah S. Narasimhan, Model identification and error covariance matrix estimation from noisy data using PCA[J]. Control Engineering Practice 2008. 16(1): p. 146-155.

[21] Teuvo Kohonen, Self-organizing maps. Vol. 30. 2001: Springer.

[22] Andrew P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms[J]. Pattern Recognition, 1997. 30(7): p. 1145-1159.

[23] Feng Aimin, Structure Driven for One-Class Classifiers Design and Extended Research [D]. Nanjing University of Aeronautics and Astronautics, Nanjing