


Techniques for Selecting the Optimal Parameters of One-Class Support Vector Machine Classifier for Reduced Samples

Yasmine Guerbai, Engineering of Electrical Systems Department, University of M'Hamed Bougara Boumerdes, Algeria

Youcef Chibani, Laboratoire d'Ingénierie des Systèmes Intelligents et Communicants, FEI, USTHB, Algeria

Yassine Meraihi, Systems Engineering and Telecommunications Laboratory, University of M'Hamed Bougara Boumerdes, Algeria

 <https://orcid.org/0000-0002-3735-7797>

ABSTRACT

Usually, the one-class support vector machine (OC-SVM) requires a large dataset for modeling effectively the target class independently to other classes. For finding the OC-SVM model, the available dataset is subdivided into two subsets, namely training and validation, which are used for training and validating the optimal parameters. This approach is effective when a large dataset is available. However, when training samples are reduced, parameters of the OC-SVM are difficult to find in absence of the validation subset. Hence, this paper proposes various techniques for selecting the optimal parameters using only a training subset. The experimental evaluation conducted on several real-world benchmarks proves the effective use of the new selection parameter techniques for validating the model of OC-SVM classifiers versus the standard validation techniques.

KEYWORDS

One-Class Support Vector Machine, Selection Parameter Technique, Training Subset, Validating Model of SVM Classifier, Validation Subset

1 INTRODUCTION

In statistics or data mining, a typical task is to learn a model from available data. The problem with evaluating such a model is that it may demonstrate adequate prediction capability on the training data, but might fail to predict future unseen data. Also, the validation protocol is the most procedure for estimating the generalization performance in this context.

Generally, cross-validation is a statistical method of evaluating and comparing the training model by dividing data into two subsets; one is used for training and the other is used for evaluating the model. In typical cross-validation, the training and validation sets must cross over in successive rounds such that each data point has a chance of being validated against. The basic form of cross-validation is k-fold cross-validation. Other forms of cross-validation are special cases of k-fold cross-validation or involve repeated rounds of k-fold cross-validation, first appeared by Mosteller and Turkey. (Mosteller & Tukey, 1968)

DOI: 10.4018/IJAMC.290533

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

In 1970, both Stone (Stone, 1974) and Geisser (Geisser, 1975) employed cross-validation. Currently, cross-validation is widely accepted in machine learning as a standard procedure for performance evaluation. There are two possible goals in cross-validation. The first goal is to estimate the performance of the learned model from available data using one algorithm. The second goal is to compare the performance of two or more different algorithms and find out the best algorithm for the available data, or alternatively to compare the performance of two or more variants of the parameterized model.

The cross-validation technique has been used by Shao (Shao, 1993) which proposed a linear model selection. Lacerda and Carvalho (De Lacerda, De Carvalho, & Ludermir, 2002) studied a problem for finding the adjustable parameters of a learning algorithm using Genetic Algorithms. This problem is also known as the model selection problem. Some model selection techniques (e.g., cross-validation and bootstrap) are combined with the Genetic Algorithms of different ways. Those combinations explore features of the Genetics Algorithms such as the ability for handling multiple and noise objective functions. Xie et al. (Xie, et al., 2011) proposed a system for testing and validating machine learning classifiers by metamorphic testing, Edwards et al (Edwards, Zhang, Parker, & New, 2013) proposed Approximate 1-fold Cross-Validation with Least Squares Support Vector Machines (LS-SVM) and Kernel Ridge Regression, The LS-SVM, a popular SVM variant, and Kernel Ridge Regression still have several scalability issues. In particular, the $O(n^3)$ computational complexity for solving a single model, and the overall computational complexity associated with tuning hyperparameters are still major problems. They addressed these problems by introducing an $O(n \log n)$ approximate 1-fold cross-validation method that uses a multi-level circulant matrix to approximate the kernel. In addition, they prove their algorithm's computational complexity and present empirical runtimes on data sets with approximately one million data points. They also validate their approximate method's effectiveness at selecting hyperparameters on real-world and standard benchmark data sets. Their system is biased because they used only the test set to validate a model, Wang et al. (Wang, Wang, & Liu, 2012) provided an improved Gaussian mixture model based on least-squares cross-validation and Gaussian Particle Swarm Optimization (PSO) algorithm with Gaussian jump. They developed an improved Gaussian mixture model (iGMM) based on least-squares cross-validation (LSCV) and Gaussian PSO with Gaussian jump (GPSOGJ). According to least-squares cross-validation, an error measure criterion is derived which is used to evaluate the estimation error between the true density function and the estimated density function. Later, Yingchao et al. (Xiao, Wang, & Xu, 2014) studied parameter Selection of Gaussian Kernel for One-Class SVM (OC-SVM). They proposed a novel method for predicting whether a new-coming sample belongs to the target class or not which is the concept of One-Class Classification. They presented an effective tool to measure the distances from the samples to the OC-SVM enclosing surfaces. Then based on this measurement, an optimization objective function for the parameter selection is put forward. Zhang et al. (Zhang, Chen, & Zhu, 2015) proposed Image denoising based on iterative generalized cross-validation and fast translation invariant, Zhang and Yang (Zhang & Yang, 2015) provided how to apply Cross-Validation to consistently choose the best method, yielding new insights and guidance for potentially vast amount of application. In addition, they addressed several seemingly widely spread misconceptions on Cross-Validation. However, the limitation of this approach is the use of two subsets (training and validating model) needing a lot of samples, Wang et al. (Wang, Liu, Zhu, & Porikli, 2018) provided a fully automatic OC-SVM hyperparameter selection method, which needs no tuning of additional hyperparameter, based on a novel self-adaptive "data shifting" mechanism: Firstly, by efficient Edge Pattern Detection (EPD) and "negatively" shifting edge patterns along the negative direction of the estimated data density gradient. Secondly, to avoid time-consuming cross-validation and enhance robustness to noise in the given training data, a pseudo target set is generated for model validation by "positively" shifting each given target datum along the positive direction of the data density gradient.

According to the state of the art, the usual techniques require two subsets for validating the classifier model, which are called training and validation, these techniques need a large number of samples in order to subdivide data into two subsets.

However, in various applications, when reduced samples are available, a validation step is hard to be applied. In order to solve this problem, several techniques are investigated for validating the classifier model without needing the validation subset. In this purpose, the objective of this paper is to study different techniques using only a training subset in the goal to select the optimal parameter of the OC-SVM classifier model.

The remainder of the paper is organized as follows. Section 2 describes a review of the OC-SVM classifier. Section 3 presents various strategies for parameter selection of the OC-SVM classifier. The experimental and statistical results are summarized in Section 4. The last section gives a summary of the proposed technique for validating the OC-SVM model.

2 REVIEW OF OC-SVM CLASSIFIER

The one-class classification consists in describing a target set of objects and detecting which new objects resemble objects in this training set. Hence, information is only available from a single class. The objects in this class are called target objects, and so all the others are outlier objects. Several different terms have been used to refer to one-class classification, such as outlier detection, novelty detection, and concept training. (Hodge & Austin, 2004) (Tax & Duin, 1999)

Several models have been proposed in the literature for one-class classification. Most can be categorized into one of the two groups: density approaches and boundary approaches. As the name implies, a density approach uses a density method to directly estimate the density of the target objects. (Barnett & Lewis), (Parra, Deco, & Miesbach, 1996), (Tarassenko, Hayton, Cerneaz, & Brady, 1995) In the testing procedure, a new sample is classified as an outlier if its surrounding region has a probability density below a specified threshold. The problem here is that sometimes it may be impossible to estimate the complete density of the data (e.g. small sample sizes). Boundary approaches (Scholkopf, Platt, Shawe-Taylor, Smola, & Williamson, 2001) have been developed focusing solely on the boundary of the data in order to overcome this kind of difficulty. Consequently, with these approaches, estimation of the complete probability density can be avoided and it becomes possible to train from the data when the exact target density distribution is unknown. Moreover, it is sufficient that the user indicates only the boundary of the target class by using examples, and there is no need to model or sample the complete distribution. (Tax D. M., 2002)

2.1 Mathematical Formulation

The OC-SVM introduced by Schölkopf et al. (Scholkopf, Platt, Shawe-Taylor, Smola, & Williamson, 2001) algorithm relies on mapping the training samples belonging to S into a high dimensional feature space leading to find the maximal margin hyperplane, which best separates the training samples from the origin. For a test pattern x , a decision function namely $f(x)$ is performed for evaluating in which side of the hyperplane it falls with $R_x = \{x \in R^d \mid f(x) > 0\}$ with a minimum volume where d is the size of the feature vector. (Rabaoui, Davy, Rossignol, & Ellouze, 2008)

Given a training set without any class information, $x_i \in R^d, i = 1, \dots, N$, the primal form of the one-class SVM is as follows:

$$\left\{ \begin{array}{l} \min_{w, \xi, \rho} \frac{1}{2} w^T w - \rho + \frac{1}{\nu l} \sum_{i=1}^m \xi_i \\ \text{s.t. } (w^T \phi(x_i)) \geq \rho - \xi_i \\ \xi_i \geq 0 \quad i = 1, \dots, l \end{array} \right. \quad (1)$$

$\phi(x_i)$ defines the mapping function while w and ρ are the solutions of the linear decision function namely $f_{oc}(x)$ taking the following form (Camci & Chinnam, 2008):

$$f_{oc}(x) = \left\{ \sum_{i=1}^n \alpha_i K(x, x_i) - \rho \right\} \quad (2)$$

Where $\phi_{ij} = K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$

The dual problem of the one-class SVM is as follows:

$$\left\{ \begin{array}{l} \min_{\alpha} \frac{1}{2} \alpha^T Q \alpha \\ \text{s.t. } 0 \leq \alpha_i \leq \frac{1}{\nu m} \quad i = 1, \dots, m \\ e^T \alpha = 1 \end{array} \right. \quad (3)$$

n is the number of training data and α_i are the Lagrange multipliers computed by optimizing the following equation:

$$\min_{\alpha} \left\{ \frac{1}{2} \alpha_i \alpha_j K(x_i, x_j) \right\} \quad (4)$$

Formally, ρ defines the distance of the hypersphere from the origin. ν is the percentage of samples considered as outliers. $K(\cdot, \cdot)$ defines the OC-SVM kernel that allows projecting samples from the original space to the feature space. This projection allows separating a sample from the origin by means of the decision function. A pattern x is then accepted when $f_{oc}(x) > 0$. Otherwise, it is rejected. Then, the solution α of equation (4) is used to compute w and ρ in Equation (1).

Figure 1 presents a simple example of the one-class SVM to illustrate how the data can be separated from the outliers. Figure 1.a shows the distribution of the genuine data (the support is indicated by the circle that encloses the data) and a small group of imposters (in this case, the outliers) in the input space. An outlier is any data instance that lies outside the support of training data. After using a suitable kernel to project the data onto the feature space, the data distribution is shown in Figure 1.b. The hyperplane w separates the training data from the origin by a maximal margin $\frac{\rho}{\|w\|}$.

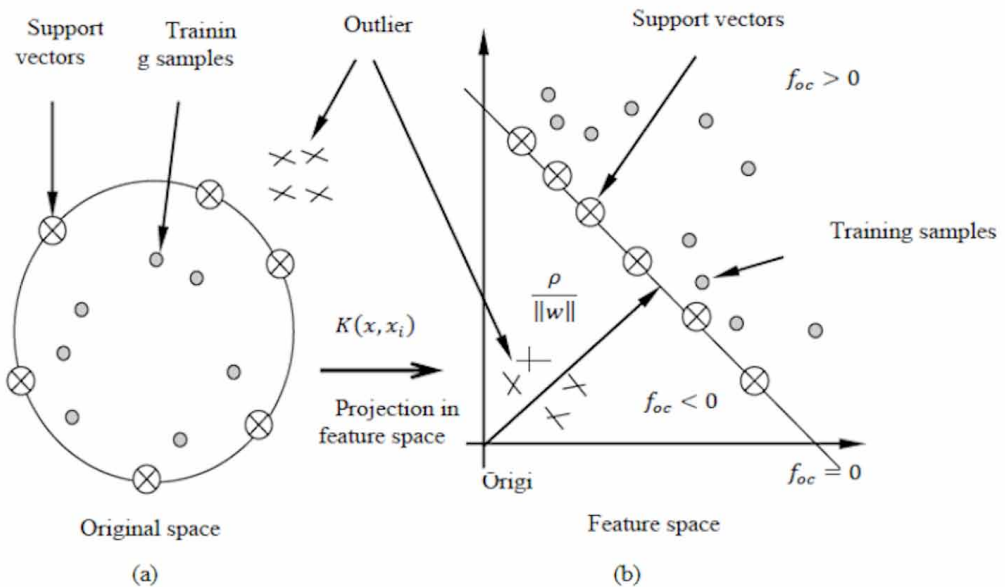
Data mapped to the same side of the origin will be given a negative one-class SVM value ($f_{oc} < 0$), whereas those mapped to the side of the training data will have positive values. In this work, the RBF kernel is used for its performance in various applications, which is defined:

$$K(x, x_i) = \exp(-\gamma d(x, x_i)) \quad (5)$$

Such that $d(x, x_i)$ is the distance between the pattern x and the training sample x_i , whereas γ is the kernel parameter.

However, unlike the traditional two-class formulation, there are no explicit penalties for false positives. Consequently, larger values of γ in the RBF kernel are required to achieve tight approximations for the performance region. It is worth noting that the SVM tends to degenerate into Parzen window estimators when a larger value for γ is used.

Figure 1. Sample classification based on OC-SVM



2.2 Extension of the OC-SVM Classifier to Multi-Class Classification

The OC-SVM was originally designed for one-class classification. (Bergamini, Oliveira, Koerich, & Sabourin, 2009) Its extension to multi-class implementation is currently an active research. (Scholkopf, Platt, Shawe-Taylor, Smola, & Williamson, 2001) The usual approach consists to represent each sample set of an object or class by its corresponding decision function of the OC-SVM (Figure 2.a).

Let define S as the set containing classes $\{\theta_1, \dots, \theta_i, \theta_M\}$ such that M is the number of classes. Hence, M OC-SVM models are built, each one is trained separately with the corresponding samples independently of other classes. A selection rule is then performed on M decision functions in order to select the best one (Figure 2.b). The straightforward way to assign a pattern x to predefined class

θ_i is to select the maximum of the decision function from the origin according to the following selection rule:

$$x \in \theta_i \text{ if } f_i(x) = \text{Max} \{f_m(x), 1 \leq m \leq M\}. \quad (6)$$

Such that

$$f_i(x) = \sum_{j=1}^{n_i} \alpha_j K(x, x_j) - \rho_i \quad (7)$$

Where n_i and ρ_i are the respective number of training samples and the distance from the origin to each OC-SVM classifier. The framework of the multi-class classification based on the OC-SVM

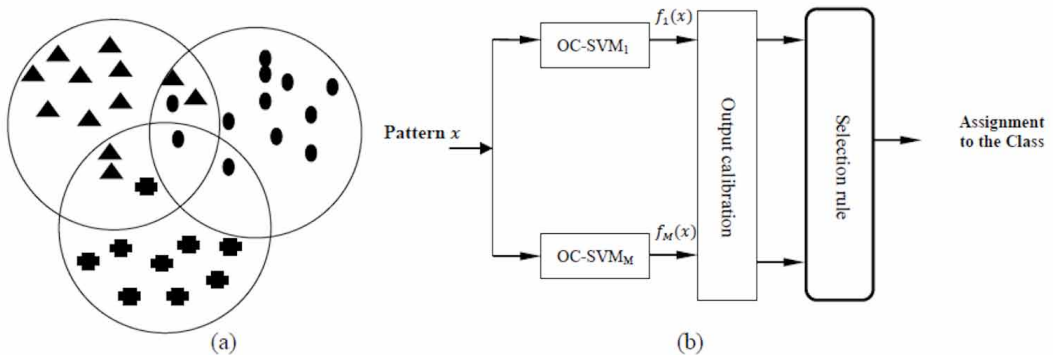
The main problem of this maximization is related to the uncalibrated outputs. Therefore, Rabaoui et al. (Rabaoui, Davy, Rossignol, & Ellouze, 2008) propose to transform the outputs using the logarithm transformation, which becomes a minimization problem. The logarithm function has the ability to attenuate the high values and increase small values for calibrating correctly the outputs of the OC-SVMs. The maximization of the decision functions becomes a minimization as follows:

$$f_i(x) = -\log \left[\sum_{j=1}^{n_i} \alpha_j K(x, x_j) \right] + \log(\rho_i) \quad (8)$$

The decision rule then takes the following equation:

$$x \in \theta_i \text{ if } f_i(x) = \text{Min} \{f_m(x), 1 \leq m \leq M\} \quad (9)$$

Figure 2. The framework of the multi-class classification based on the OC-SVM



3 PARAMETER SELECTION OF THE OC-SVM CLASSIFIER

Usually, the OC-SVM model is validated using training and validation subsets. However, when reduced samples are available, a validation step is hard to be performed. In order to overcome this problem, this paper proposes to investigate several techniques for selecting the optimal parameters in order to validate the OC-SVM model without using the validation subset which is our objective.

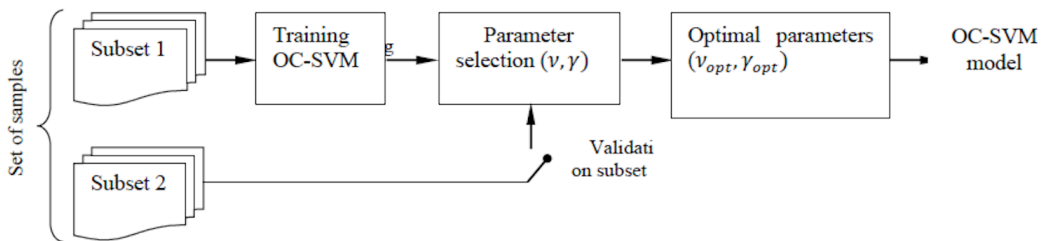
For selecting the optimal parameters, two cases are considered: the first one is the novelty detection case and the second one is the multi-class classification case.

The OC-SVM has two parameters (ν, γ) , which are fixed through the training and validation steps. In the training step, different couples of parameters are generated in order to achieve the best recognition rate of the training dataset. Usually, two subsets are used for validating the OC-SVM model, which are the training and the validation subsets, respectively. Hence, in this paper, other techniques are investigated for validating the OC-SVM model without using the validation subset. It means that the model of OC-SVM classifier can be validated using only a training subset.

The experimental protocol is then conducted into two steps. The first one consists to use jointly the training and validation samples for validating the OC-SVM model which is the standard technique. The second one that is our objective consists to validate the OC-SVM model using only the training samples.

Hence, the set of samples for each class is divided into two subsets namely Subset 1 containing M_t samples and Subset 2 containing M_v samples, respectively. Figure 3 shows the concept for selecting the optimal parameters of the OC-SVM classifiers with and without the validation subset.

Figure 3. The concept for selecting the optimal parameters of the OC-SVM classifier with and without the validation samples



The standard technique for validating the OC-SVM classifier model consists to use two subsets, the subset 1 allows to train the OC-SVM classifier and generating several couples of (ν, γ) whereas the subset 2 is used for selecting the optimal parameter using the maximum of recognition rate.

The proposed approach for validating the model consists to use only one subset which is the subset 1 to validate the OC-SVM model by using different techniques. These latter are based on the calculation of the maximum rate during the training step, the maximum number of the support vectors and the maximum number of the decision functions.

From the decision function, three main rules can be used for validating the OC-SVM model, which are

- The recognition rate during the training step.

The recognition rate during the training step (Max_RT) allows estimating sampling variability, which enables to determine whether the data is well-tuned.

- The maximum amplitude of decision function.

As mentioned in section 2, the maximum amplitude of the decision function can be deduced using equation 2. ($\#Max_D$).

- The maximum number of support vectors.

The decision function of SVM can be supported only using the exterior patterns in each data class, which are called support vectors. Motivated by this, it is useful to calculate the maximum number of supports $\#Max_SV$ in order to validate the model of OC-SVM classifier.

In summary, the proposed validation techniques are performed according to Algorithm 1 and 2, respectively.

Algorithm 1: Validating OC-SVM using validation samples

From the Subset 1 and 2, parameters of OC-SVM classifier are computed according to the following steps:
1) Training the OC-SVM classifier using training samples.
2) Validating the model of OC-SVM classifier using the maximum recognition of validation samples of Subset 2.
3) Testing using test samples according to the equation 2 in the case of novelty detection and equation 9 in the case of multi-class classification.

Algorithm 2: Validating OC-SVM without using validation samples

From the Subset 1, parameters of OC-SVM classifier are computed according to the following steps:
1) Training the OC-SVM classifier using training samples.
2) Validating the model of OC-SVM classifier using the maximum rate during the training step, the maximum number of the support vectors and the maximum number of the decision functions using training samples.
3) Testing using test samples according to the equation 2 in the case of novelty detection and equation 9 in the case of multi-class classification.

4 EXPERIMENTAL METHODOLOGY

4.1 Dataset Description

In order to evaluate the performance of the validation techniques of OC-SVM classifier, real-world benchmarks are used for selecting the optimal parameters of OC-SVM classifiers considering two cases: novelty detection and multi-class classification. Hence, seven different datasets from UCI datasets (Frank, 2010) are used and handwritten digits datasets, (Hull, 1994) which are reported in Table 1. Each dataset is randomly divided into three subsets for training and validating the classifier while the test dataset is used for evaluating its performance. $\#$ Classes and $\#$ Features are the numbers of classes and the size of the feature vectors. $\#$ Training, $\#$ Validation and $\#$ Test samples of the number samples used for training, validating and testing the OC-SVM classifiers. For evaluating the various validation techniques, LibSVM library is used for SVM classification through the Matlab software.

Table 1. Datasets used for evaluating the proposed validation techniques

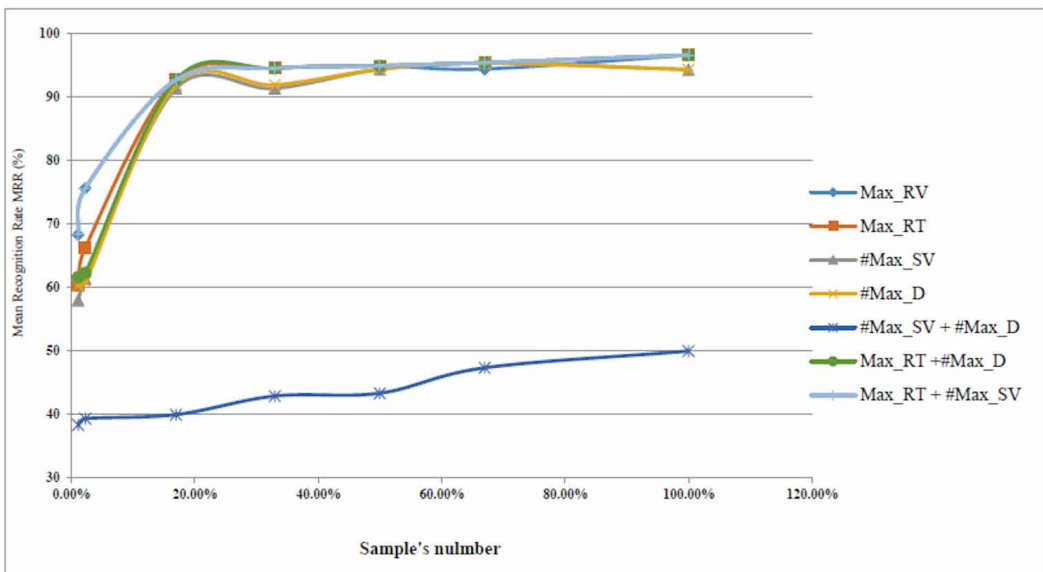
Dataset	# Classes	# Features	# Training samples	# Validation samples	# Test samples
Banana	2	2	60	20	120
Glass	2	9	100	20	94
Pima	2	8	350	50	368
Breast Cancer	2	9	350	50	299
Iris	3	4	45	45	60
Satimage	6	36	600	600	5235
Digits (USPS)	10	14	3646	3641	2007
Texture	11	40	732	731	4037

4.2 Experimental Protocol

The comparative evaluation starts by applying the standard approach for validating the OC-SVM by considering the maximum rate during the validation step (Max_RV). Then, various techniques are proposed based on the selection of the optimal parameters without needing to use a validation step. These techniques are based on the calculation of the maximum rate during the training step (Max_RT), the maximum number of the support vectors (#Max_SV), the maximum number of the decision functions (#Max_D). Moreover, these techniques are conjointly used by considering (#Max_SV + #Max_D), (Max_RT + #Max_D) and (Max_RT + #Max_SV).

The proposed different techniques are evaluated on novelty detection and multi-class classification using Mean Recognition Rate (MRR).

Figure 4. Influence of sample's number for Breast Cancer dataset



4.3 Validation Techniques for Novelty Detection

For novelty detection, four datasets including two classes are considered. Hence, only OC-SVM is trained in one-class. In order to study the influence of the sample's number used for training the OC-SVM, samples are selected from 4 to the whole dataset expressed in %. Results presented for each dataset in terms of MRR are reported in figures 4, 5, 6, and 7, respectively.

Figure 5. Influence of sample's number for Glass dataset

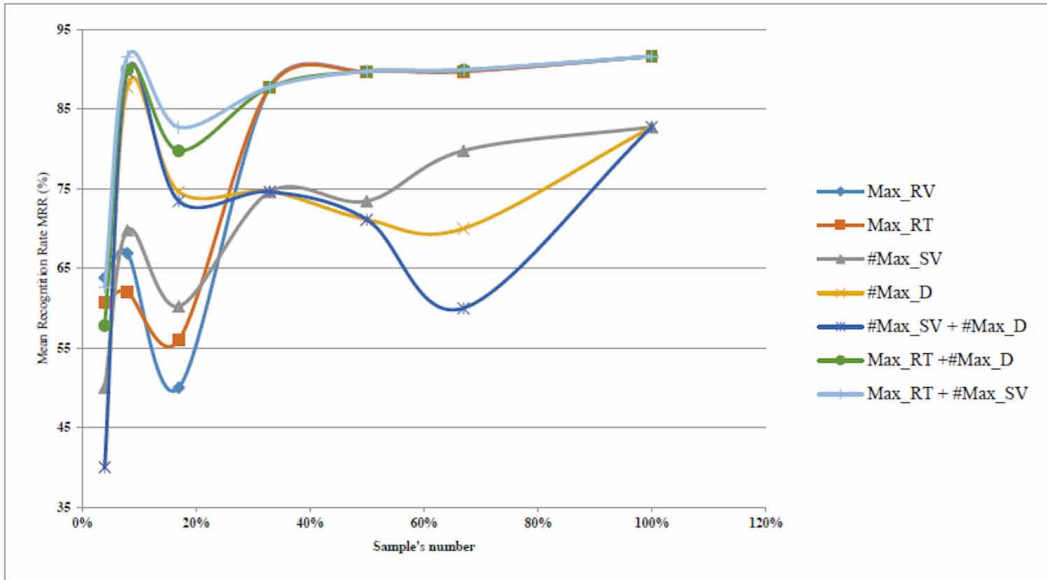


Figure 6. Influence of sample's number for Pima Indian diabetes dataset

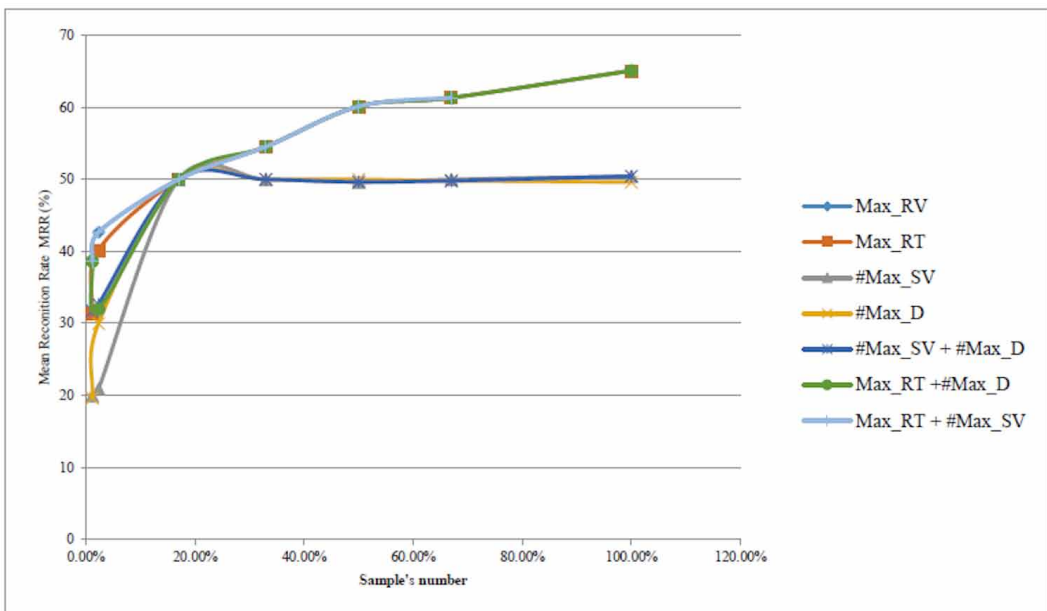
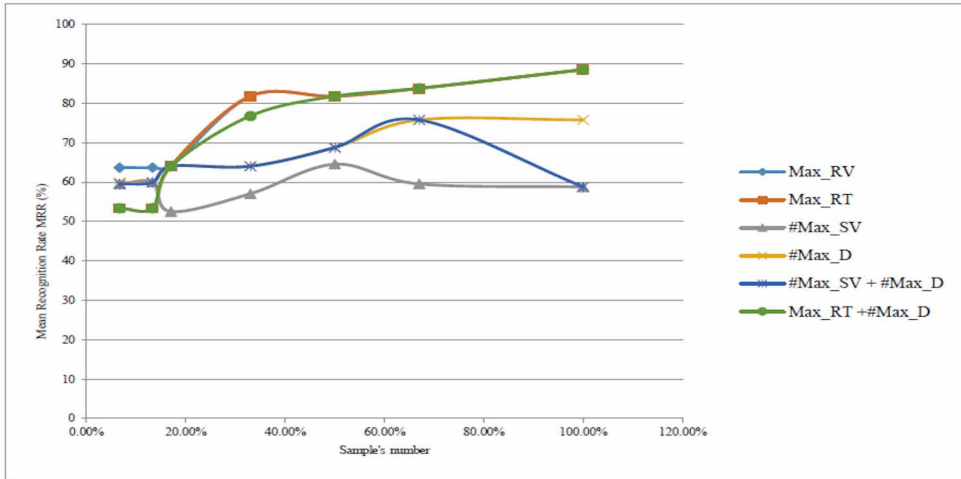


Figure 7. Influence of sample's number for Banana dataset



From Tables 2 and 3, it is straightforward to observe that when few samples are used for training, results obtained for Max_RV is comparable to Max_RT combined with #Max_SV (Max_RT + #Max_SV). It is also seen that the validation technique based on Max_RT + #Max_SV is more stable compared to other techniques. Therefore, it is suitable and recommended for its use for novelty detection when few samples are available.

Table 2. Novelty detection accuracy using reduced samples from the training dataset

Dataset	With validation	Without validation					
	Max_RV	Max_RT	#Max_SV	#Max_D	#Max_SV + #Max_D	Max_RT + #Max_D	Max_RT + #Max_SV
Breast cancer	68.39	60.37	58.00	60.37	38.37	61.66	68.39
Glass	65.35	63.99	63.99	35.38	35.38	60.69	65.35
Pima Indian diabetes	40.66	38.04	20.00	20.00	28.00	38.08	40.00
Banana	62.58	60.08	35.66	57.00	57.33	57.33	62.58

Table 3. Novelty detection accuracy using large samples from the training dataset

Dataset	With validation	Without validation					
	Max_RV	Max_RT	#Max_SV	#Max_D	#Max_SV + #Max_D	Max_RT + #Max_D	Max_RT + #Max_SV
Breast cancer	96.39	96.39	96.19	96.19	96.19	96.39	96.39
Glass	90.69	90.69	55.58	25.58	55.58	90.69	90.69
Pima Indian diabetes	57.04	57.04	28.00	28.57	28.57	57.04	57.04
Banana	87.50	87.50	55.00	77.50	55.00	87.50	87.50

4.4 Validation Techniques for Multi-Class Classification

In order to evaluate the effective use of the proposed techniques for multi-class classification, Tables 4 and 5 report MRR obtained for seven datasets (Breast cancer, Glass, Pima Indian- diabetes and banana, Iris, Satimag, Digits (USPS) and texture) when selecting reduced and large samples from training dataset, respectively. It is observed in Table 4 that the joint use of Max_RT and #Max_SV (Max_RT+#Max_SV) are more suitable when reduced samples are available since similar MRR are obtained when using the validation rule (Max_RV). In contrast, Table 5 shows that Max_RT or Max_RT used jointly with #Max_D or #Max_SV provide similar MRR when large samples are selected from the training dataset. Therefore, when a few samples are available, the joint use of Max_RT and #Max_SV is more suitable.

Table 4. Multi-class classification accuracy using reduced samples from the training dataset

Dataset	With validation	Without validation					
	Max_RV	Max_RT	#Max_SV	#Max_D	#Max_SV + #Max_D	Max_RT + #Max_D	Max_RT + #Max_SV
Breast cancer	68.39	60.37	58.00	60.37	38.37	61.66	68.39
Glass	65.35	63.99	63.99	35.38	35.38	60.69	65.35
Pima Indian diabetes	40.66	38.04	20.00	20.00	28.00	38.08	40.00
Banana	62.58	60.08	35.66	57.00	57.33	57.33	62.58
Iris	69.66	69.66	60.69	57.89	57.89	60.00	69.66
Satimag	60.16	57.66	57.66	57.66	59.33	59.16	60.16
Digits (USPS)	71.36	63.12	50.00	50.00	66.40	66.40	71.36
Texture	75.33	70.66	60.85	70.33	70.00	70.66	75.33

Table 5. Multi-class classification accuracy using large samples from the training dataset

Dataset	With validation	Without validation					
	Max_RV	Max_RT	#Max_SV	#Max_D	#Max_SV + #Max_D	Max_RT + #Max_D	Max_RT + #Max_SV
Breast cancer	96.39	96.39	96.19	96.19	96.19	96.39	96.39
Glass	90.69	90.69	55.58	25.58	55.58	90.69	90.69
Pima Indian diabetes	57.04	57.04	28.00	28.57	28.57	57.04	57.04
Banana	87.50	87.50	55.00	77.50	55.00	87.50	87.50
Iris	89.66	89.66	80.00	78.00	80.00	89.66	89.66
Satimag	78.33	78.33	72.66	72.66	73.66	78.33	78.33
Digits (USPS)	93.12	93.12	86.40	85.30	86.40	93.12	93.12
Texture	92.89	92.89	90.85	89.33	90.85	92.89	92.89

5 CONCLUSION

The objective of this paper aimed to propose various techniques for validating the OC-SVM model without needing the validation subset. Experimental results conducted on several real-world datasets show that the proposed techniques achieve similar results like the validation technique according to the selected training dataset. Indeed, the technique using the maximum of training (Max_RT) associated with the maximum number of support vectors (#Max_SV) is suitable when the reduced number of training samples is available whatever the type of classification (novelty detection or multi-class classification). In contrast, when large samples are available, a simple technique can be used based on the maximum rate. Therefore, the choice of the validation technique of the OC-SVM model depends on the available number of training samples (Max_RT). According to applications, the Max_RT rule can be used when large samples are available. In contrast, the validation technique based on Max_RT associated with #Max_SV is used when a reduced number of selected training samples is applied.

REFERENCES

- Barnett, V., & Lewis, T. (1974). *Outliers in statistical data*. Wiley.
- Bergamini, C., Oliveira, L. S., Koerich, A. L., & Sabourin, R. (2009). Combining different biometric traits with one-class classification. *Signal Processing*, 89(11), 2117–2127. doi:10.1016/j.sigpro.2009.04.043
- Camci, F., & Chinnam, R. B. (2008). General support vector representation machine for one-class classification of non-stationary classes. *Pattern Recognition*, 41(10), 3021–3034. doi:10.1016/j.patcog.2008.04.001
- De Lacerda, E. G., De Carvalho, A., & Ludermir, T. B. (2002). A study of cross-validation and bootstrap as objective functions for genetic algorithms. *VII IEEE Brazilian Symposium on Neural Networks*.
- Edwards, R. E., Zhang, H., Parker, L. E., & New, J. R. (2013). Approximate l-fold cross-validation with least squares SVM and kernel ridge regression. *12th International Conference on Machine Learning and Applications*, 1, 58-64. doi:10.1109/ICMLA.2013.18
- Frank, A. (2010). *Uci machine learning repository*. University of California, School of Information and Computer Science. <http://archive.ics.uci.edu/ml>
- Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70(350), 320–328. doi:10.1080/01621459.1975.10479865
- Hodge, V., & Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2), 85–126. doi:10.1023/B:AIRE.0000045502.10941.a9
- Hull, J. J. (1994). A database for handwritten text recognition research. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5), 550–554. doi:10.1109/34.291440
- Mosteller, F., & Tukey, J. W. (1968). Data analysis, including statistics. *Handbook of social psychology*, 2, 80-203.
- Parra, L., Deco, G. D., & Miesbach, S. (1996). Statistical independence and novelty detection with information preserving nonlinear maps. *Neural Computation*, 8(2), 260–269. doi:10.1162/neco.1996.8.2.260
- Rabaoui, A., Davy, M., Rossignol, S., & Ellouze, N. (2008). Using one-class svms and wavelets for audio surveillance. *IEEE Transactions on Information Forensics and Security*, 3(4), 763–775. doi:10.1109/TIFS.2008.2008216
- Scholkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7), 1443–1471. doi:10.1162/089976601750264965 PMID:11440593
- Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association*, 88(422), 486–494. doi:10.1080/01621459.1993.10476299
- Stone, M. (1974). Cross-validated choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B. Methodological*, 36(2), 111–133. doi:10.1111/j.2517-6161.1974.tb00994.x
- Tarassenko, L., Hayton, P., Cerneaz, N., & Brady, M. (1995). *Novelty detection for the identification of masses in mammograms*. Academic Press.
- Tax, D. M. (2002). *One-class classification: Concept learning in the absence of counter-examples*. Academic Press.
- Tax, D. M., & Duin, R. P. (1999). Support vector domain description. *Pattern Recognition Letters*, 20(11-13), 1191–1199. doi:10.1016/S0167-8655(99)00087-2
- Wang, S., Liu, Q., Zhu, E., Porikli, F., & Yin, J. (2018). Hyperparameter selection of one-class support vector machine by self-adaptive data shifting. *Pattern Recognition*, 74, 198–211. doi:10.1016/j.patcog.2017.09.012
- Wang, X., Wang, H. B., & Liu, H. T. (2012). An improved Gaussian mixture model based on least-squares cross-validation and Gaussian PSO with Gaussian jump. *International Conference on Machine Learning and Cybernetics*, 2, 714-719. doi:10.1109/ICMLC.2012.6359012

- Xiao, Y., Wang, H., & Xu, W. (2014). Parameter selection of Gaussian kernel for one-class SVM. *IEEE Transactions on Cybernetics*, 45(5), 941–953. doi:10.1109/TCYB.2014.2340433 PMID:25099969
- Xie, X., Ho, J. W. K., Murphy, C., Kaiser, G., Xu, B., & Chen, T. Y. (2011). Testing and validating machine learning classifiers by metamorphic testing. *Journal of Systems and Software*, 84(4), 544–558. doi:10.1016/j.jss.2010.11.920 PMID:21532969
- Zhang, L., Chen, J., & Zhu, T. (2015). Image denoising based on iterative generalized cross-validation and fast translation invariant. *Journal of Visual Communication and Image Representation*, 28, 1–14. doi:10.1016/j.jvcir.2015.01.002
- Zhang, Y., & Yang, Y. (2015). Cross-validation for selecting a model selection procedure. *Journal of Econometrics*, 187(1), 95–112. doi:10.1016/j.jeconom.2015.02.006

Yasmine Guerbai received her Master's degree in intelligent and communicating system from the University of Science and Technology Houari Boumedienne, Algiers, Algeria in 2010 and received her PhD from the University of Science and Technology Houari Boumedienne in 2016. Her main current research interests include machine learning, pattern recognition, classification, feature generation, identification and verification of writers, image and signal processing.

Youcef Chibani was born in Algiers and is currently working as a full professor. His research interest includes image and signal processing, machine learning and their applications for document analysis including writer identification, signature verification and historical document analysis. He was the head of the Scientific Council, responsible for master and doctoral training, director of the Communicating and Intelligent System Engineering Laboratory and currently is the Dean of the Faculty of Electronics and Computer Science. He co-authored various book chapters and papers in international peer-reviewed journal and conferences.

Yassine Meraihi received his Master's degree in Computer Sciences from the High School of Computer Sciences, E.S.I, ex INI, Algeria in 2003 and received his PhD from the University of MHamed Bougara Boumerdes, Algeria in 2017. He is currently an Associate Professor at the University of MHamed Bougara Boumerdes, Algeria. His main current research interests include quality of service for multimedia wired and wireless networks, routing in challenged networks including WMSNs and VANETs and applications of metaheuristics to optimization problems.