IMPROVED ONE-CLASS SVM CLASSIFIER FOR SOUNDS CLASSIFICATION

A. Rabaoui[1] M. Davy[2] S. Rossignol[2] Z. Lachiri[1] N. Ellouze[1]
[1]Unité de recherche Signal, Image et Reconnaissance des formes, ENIT, BP 37, Campus Universitaire, 1002 le Belvédère, Tunis Tunisia
[2]LAGIS, UMR CNRS 8146, and INRIA SequeL Team, BP 48, Cité Scientifique, 59651 Villeneuve d'Ascq Cedex, Lille France

Abstract

This paper proposes to apply optimized One-Class Support Vector Machines (1-SVMs) as a discriminative framework in order to address a specific audio classification problem. First, since SVM-based classifier with gaussian RBF kernel is sensitive to the kernel width, the width will be scaled in a distribution-dependent way permitting to avoid under-fitting and over-fitting problems. Moreover, an advanced dissimilarity measure will be introduced. We illustrate the performance of these methods on an audio database containing environmental sounds that may be of great importance for surveillance and security applications. The experiments conducted on a multi-class problem show that by choosing adequately the SVM parameters, we can efficiently address a sounds classification problem characterized by complex real-world datasets.

1 Introduction

Support Vector Machine (SVM) classifiers have been shown to provide better performance than more traditional techniques in many problems, thanks to their ability to generalize. In many cases, SVMs outperform most state-of-the-art classifiers. In recent years, many pattern recognition problems have been tackled using SVMs, ranging from computer vision to text classification. However, their application to computer auditory scene analysis (CASA) and, more specifically to sound recognition problems, has been very limited.

In this paper we focus on the specific task of impulsive sounds classification for a surveillance application using several one class SVMs (1-SVMs) [22]. Though they have been less studied than two-class SVMs (2-SVMs), 1-SVMs have proved extremely powerful in audio applications [22, 10, 9]. Moreover, we found that the 1-SVM methodology is more easy to implement and tune, while having a reasonable training cost. The application addressed here concerns sound classification. In real environment, there might be many sounds which do not belong to one of the pre-defined classes, thus it is necessary to define a *rejection class*, which may gather all sounds which do not belong to the training classes. An easy and elegant way to do so consists of estimating the regions of high probability of the known classes in the space of features, and considering the rest of the space as the rejection class. Training several 1-SVMs does this automatically.

As many others, the 1-SVM approach is dependent on a set of user-tuned parameters which have a regularization effect during training. Also, since these parameters are not learnt, they may be chosen inadequatly and lead to poor performance on a given data set. Thus, a systematic methodology that allows for parameter optimization is quite useful. Our investigations will show that one effective method for improving 1-SVMs consists of well selecting the 1-SVM kernel function, and thus improving sound class prediction.

In the literature, several attempts have been made to improve class-prediction accuracy of SVMs [3, 19, 16, 29]. In [1, 2], a conformal transformation method was proposed to change the kernel function. Another method is proposed in [15] which is based on the kernel-alignment idea but with a simple transformation of the ideal target kernel, to adapt the kernel in the imbalanced training-data problem. Compared to [15], our method deals with the entire training dataset and not just the class-boundary data. Moreover, this solution modify the kernel function instead of the kernel matrix as in [15].

In this paper, we assume that the kernel function is a Radial Basis (RBF) function and we follow [8] to deal with the optimal σ which is the tradeoff between the over-fitting loss in dense areas and the under-fitting loss in sparse areas, by scaling the kernel width in a distribution-dependent way. In addition, we will introduce an advanced decision function formulation allowing for a sophisticated dissimilarity measure.

The remainder of this paper is organized as follows. Section 2 gives an overview of the 1-SVM-based sound classi-

978-1-4244-1696-7/07/$25.00 ©2007 IEEE.

fier: 1-SVM theory and derivation of an advanced dissimilarity measure. Section 3 discusses the kernel scaling width method. Experimental set-up and results are provided in Section 4. Section 5 concludes the paper with a summary.

2 One-class SVM classifier

The 1-SVM approach[1] [27] has been successfully applied to various learning problems [9, 20, 6, 28, 13]. It consists of learning the minimum volume contour that encloses most of the data in a dataset. Its original application is outlier detection, to detect data that differ from most of the data within a dataset.

More precisely, let $\mathcal{X} = \{x_1, \ldots, x_m\}$ a dataset in $\mathbb{R}^d$. Here, each x_j is the full feature vector of a signal, i.e., each signal is represented by one vector x_j in $\mathbb{R}^d$. The aim of 1-SVMs is to use the training data so as to learn a function $\mathsf{f}_{\mathcal{X}} : \mathbb{R}^d \mapsto \mathbb{R}$ such that most of the data in $\mathcal{X}$ belong to the set $\mathcal{R}_{\mathcal{X}} = \{x \in \mathbb{R}^d \text{ with } \mathsf{f}_{\mathcal{X}}(x) \geq 0\}$ while the volume of $\mathcal{R}_{\mathcal{X}}$ is minimal. This problem is termed *minimum volume set (MVS) estimation*, see [11], and we see that membership of x to $\mathcal{R}_{\mathcal{X}}$ indicates whether this datum is overall similar to $\mathcal{X}$, or not. Thus, by learning regions $\mathcal{R}_{\mathcal{X}_i}$ for each class of sound $(i = 1, \ldots, N)$, we learn N membership functions $\mathsf{f}_{\mathcal{X}_i}$. Given the $\mathsf{f}_{\mathcal{X}_i}$'s, the assignment of a datum x to a class is performed as detailed in Section 2.1.

1-SVMs solve MVS estimation in the following way. First, a so-called *kernel function* $k(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$ is selected, and it is assumed *positive definite*, see [27]. Here, we assume a Gaussian RBF kernel such that $k(x,x') = \exp\big[-\|x-x'\|^2/2\sigma^2\big]$, where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^d$. This kernel induces a so-called *feature space* denoted $\mathcal{H}$ via the mapping $\phi : \mathbb{R}^d \mapsto \mathcal{H}$ defined by $\phi(x) \triangleq k(x,\cdot)$, where $\mathcal{H}$ is shown to be reproducing kernel Hilbert space (RKHS) of functions, with dot product denoted $\langle\cdot,\cdot\rangle_{\mathcal{H}}$. The reproducing kernel property implies that $\langle\phi(x),\phi(x')\rangle_{\mathcal{H}} = \langle k(x,\cdot),k(x',\cdot)\rangle_{\mathcal{H}} = k(x,x')$ which makes the evaluation of $k(x,x')$ a linear operation in $\mathcal{H}$, whereas it is a nonlinear operation in $\mathbb{R}^d$. In the case of the Gaussian RBF kernel, we see that $\|\phi(x)\|^2_{\mathcal{H}} \triangleq \langle\phi(x),\phi(x)\rangle_{\mathcal{H}} = k(x,x) = 1$, thus all the mapped data are located on the hypersphere with radius one, centered onto the origin of $\mathcal{H}$ denoted $S_{(o,R=1)}$, see Fig 1. The 1-SVM approach proceeds in $\mathcal{H}$ by determining the hyperplane $\mathcal{W}$ that separates most of the data from the hypersphere origin, while being as far as possible from it. Since in $\mathcal{H}$, the image by ϕ of $\mathcal{R}_{\mathcal{X}}$ is included in the segment of hypersphere bounded by $\mathcal{W}$, this indeed implements MVS estimation [11]. In practice, let $\mathcal{W} = \{h(\cdot) \in \mathcal{H} \text{ with } \langle h(\cdot), w(\cdot)\rangle_{\mathcal{H}} - \rho = 0\}$, then its parameters $w(\cdot)$ and ρ results from the optimization problem

$$\min_{\mathrm{w},\xi,\rho} \frac{1}{2}\|w(\cdot)\|^2_{\mathcal{H}} + \frac{1}{\nu m}\sum_{j=1}^{m}\xi_j - \rho \qquad (1)$$

subject to (for $i = 1, \ldots, m$)

$$\langle w(\cdot), k(x_j,\cdot)\rangle_{\mathcal{H}} \geq \rho - \xi_j, \quad \text{and } \xi_j \geq 0 \qquad (2)$$

where ν tunes the fraction of data that are allowed to be on the wrong side of $\mathcal{W}$ (these are the outliers and they do not belong to $\mathcal{R}_{\mathcal{X}}$) and ξ_j's are so-called slack variables. It can be shown [27] that a solution of (1)-(2) is such that

$$w(\cdot) \;=\; \sum_{j=1}^{m}\alpha_j k(x_j,\cdot) \qquad (3)$$

where the α_j's verify the dual optimization problem

$$\min_{\alpha} \frac{1}{2}\sum_{j,j'=1}^{m}\alpha_j\alpha_{j'}k(x_j,x_{j'}) \qquad (4)$$

subject to $0 \leq \alpha_j \leq \frac{1}{\nu m}$, $\sum_j \alpha_j = 1$. Finally, the decision function is $\mathsf{f}_{\mathcal{X}}(x) = \sum_{j=1}^{m}\alpha_j k(x_j,x) - \rho$ and ρ is computed by using that $\mathsf{f}_{\mathcal{X}}(x_j) = 0$ for those x_j's in $\mathcal{X}$ that are located onto the boundary, i.e., those that verify both $\alpha_j \neq 0$ and $\alpha_j \neq 1/\nu m$. An important remark is that the solution is sparse, i.e., most of the α_i's are zero (they correspond to the x_j's which are inside the region $\mathcal{R}_{\mathcal{X}}$, and they verify $\mathsf{f}_{\mathcal{X}}(x) > 0$).

As plotted in Fig. 1, the MVS in $\mathcal{H}$ may also be estimated by finding the minimum volume hypersphere that encloses most of the data (Support Vector Data Description (SVDD) [28]), but this approach is equivalent to the hyperplane one in the case of a RBF kernel.

In order to adjust the kernel for optimal results, the parameter σ can be tuned to control the amount of smoothing, i.e. large values of σ lead to flat decision boundaries, as shown in Fig. 2. Also, ν is an upper bound on the fraction of outliers in the dataset [27]. Fig. 2 displays an example of a 2-dimensional data to which a 1-SVM using an RBF kernel was applied for different values of ν and σ.

2.1 A dissimilarity measure

The 1-SVM described above can be used to learn the MVS of a dataset of feature vectors which relate to sounds. This is not yet a multiclass classifier, and we now describe how to build such a classifier, by adapting the results of [9, 12].

Assume that N 1-SVMs have been learnt from the datasets $\{\mathcal{X}_1, \ldots, \mathcal{X}_N\}$, and consider one of them, with associated set of coefficients denoted $(\{\alpha_j\}_{j=1,\ldots,m}, \rho)$. In

[1] Various terms have been used in the literature to refer to one-class learning approaches. The term single-class classification originates from Moya [21], but also outlier detection [25], novelty detection [10, 4] or concept learning [14] are used.

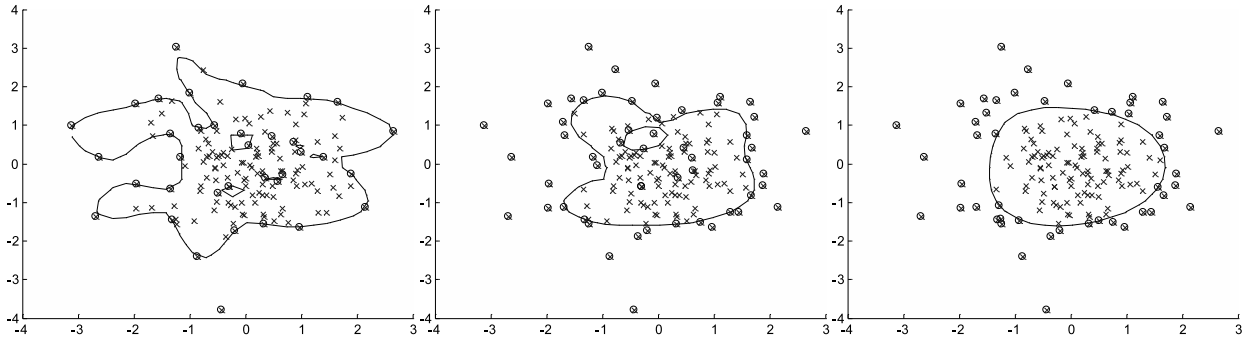


Figure 2: *A Single-class SVM using an RBF kernel is applied to 2-dimensional data; the parameter ν characterizes the fractions of* SVs *(Support Vectors) and* outliers *(points which are on the wrong side of the hyperplane). The parameter σ characterizes the kernel width. For these plots, the parameters are (a) $\nu = 0.1$, $\sigma = 1$ (b) $\nu = 0.3$, $\sigma = 1$ (c) $\nu = 0.3$, $\sigma = 4$.*

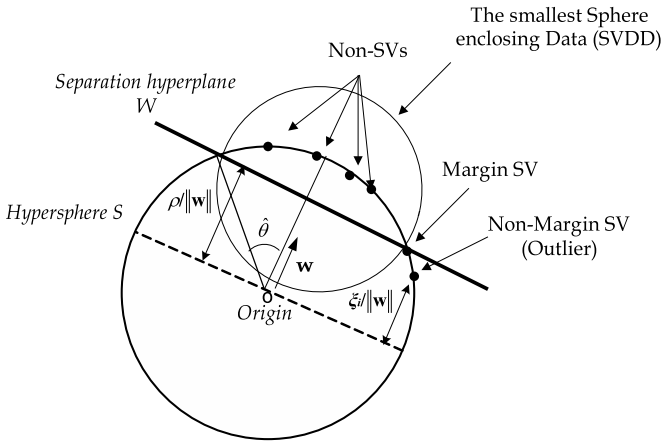


Figure 1: *In the feature space $\mathcal{H}$, the training data are mapped on a hypersphere $S_{(o,R=1)}$. The 1-SVM algorithm defines a hyperplane with equation $\mathcal{W} = \{h \in \mathcal{H}$ s.t. $\langle w, h\rangle_{\mathcal{H}} - \rho = 0\}$, orthogonal to w. Black dots represent the set of mapped data, that is, $k(\mathrm{x}_j, \cdot), i = 1, ..., m$. For RBF kernels, which depend only on $x - x'$, $k(x, x')$ is constant, and the mapped data points thus lie on a hypersphere. In this case, finding the smallest sphere enclosing the data is equivalent to maximizing the margin of separation from the origin.*

order to determine whether a new datum x is similar to the set $\mathcal{X}$, we define a dissimilarity measure, denoted $d(\mathcal{X}, x)$ as follows

$$d(\mathcal{X}, x) \;=\; -\log[\sum_{j=1}^{m} \alpha_j \, k(x, x_j)] + \log[\rho] \qquad (5)$$

in which ρ is seen as a scaling parameter which balances the α_j's. Thanks to this normalization, the comparison of such dissimilarity measures $d(\mathcal{X}_i, x)$ and $d(\mathcal{X}_{i'}, x)$ is possible. Indeed,

$$d(\mathcal{X}, x) \;=\; -\log\big[\frac{\langle w(\cdot), k(x, \cdot)\rangle_{\mathcal{H}}}{\rho}\big] \qquad (6)$$

because $\|k(x, \cdot)\|_{\mathcal{H}} = 1$, where $w(\cdot)\angle k(x, \cdot)$ denotes the angle between $w(\cdot)$ and $k(x, \cdot)$. By doing elementary geometry in $\mathcal{H}$, we can show that $\frac{\rho}{\|w(\cdot)\|_{\mathcal{H}}} = \cos(\widehat{\theta})$, see Fig.1. This yields the following interpretation of $d(\mathcal{X}, x)$

$$d(\mathcal{X}, x) \;=\; -\log\big[\frac{\cos\big(w(\cdot)\angle k(x, \cdot)\big)}{\cos(\widehat{\theta})}\big] \qquad (7)$$

which shows that the normalization is sound, and makes $d(\mathcal{X}, x)$ a valid tool to examine the membership of x to a given class represented by a training set $\mathcal{X}$.

2.2 Multiple sound classes 1-SVM-based classification Algorithm

The sound classification algorithm comprises three main steps. Step one is that of training data preparation, and it includes the selection of a set of features (MFCCs in experiments) which are computed for all the training data on sliding windows. Thus, each sound is converted into a time series of such features, which are then averaged over the signal. The resulting feature vectors[2] are centered and normalized to unit variances. The value of ν is selected in the reduced interval $[0.05, 0.8]$ in order to avoid edge effects for small or large values of ν.

[2]We emphasize the use of *feature* is confusing and that the features vectors are different from data in feature space, the former being finite dimensional vectors and the latter being functions. However, the word being used in both audio signal processing and VMs, we stick to these two uses.

Algorithm 1: Sound classification algorithm

Step 1: Data preparation

- Form the training sets $\mathcal{X}_i = \{x_{i,1}, ..., x_{i,m_i}\}$, $i = 1, \ldots, N$ by computing the feature vectors for all the training sounds selected.
- Set the parameter σ of the Gaussian RBF kernel to some pre-determined value (e.g., set σ as half the average euclidean distance between any two points $x_{i,j}$ and $x_{i',j'}$, see [26]), and select $\nu \in [0.05, 0.8]$.

Step 2: Training step

- For $i = 1, \ldots, N$, solve the 1-SVM problem for the set $\mathcal{X}_i$, resulting in a set of coefficients ($\alpha_{i,j}$, ρ_j), $j = 1, \ldots, m_i$

Step 3: Testing step

- For each sound s to be classified into one of the N classes, do
  - compute its feature vector, denoted x,
  - for $i = 1, \ldots, N$, compute $d(\mathcal{X}_i, x)$ by using Eq. (5)
  - assign the sound s to the class $\widehat{i}$ such that $\widehat{i} = \arg\min_{i=1,\ldots,N} d(\mathcal{X}_i, x)$

3. Scaling the kernel width

We assume here that the Gaussian RBF kernel is used. This provides a powerful learner, however it is sensitive to the choice of the scale (or width) parameter σ. This parameter controls the trade-off between faithfulness to the training data and smoothness of the decision surface in the space of the data. When σ is too small the SVM overfits the training data. If σ is too large then the decision boundary is not able to model the required decision boundary. In most cases, when using a Gaussian kernel the scale parameter σ must be specified in advance.

In a given classification task, picking the best kernel parameters is actually a non trivial model selection problem. It may be solved by either an exhaustive search over all parameter values or an optimization procedure that explores only a finite subset of the possible values. Generally, the parameters are selected empirically by trying a finite number of values and keeping those that minimize the classification error estimated over a validation set. This procedure requires, e.g., a grid search over the parameter space and needs to locate the interval of feasible solution. The overall process of building many models and choosing the best one may be time-consuming.

In this section, we explore an information-geometrical structure induced in the input space by the kernel. [23, 7, 17, 5] discuss the selection of σ for a Gaussian RBF kernel so as to avoid the problem of over-fitting and under-fitting in SVM learning. In the 1-SVM classification task with Gaussian RBF kernel, the kernel width must be adapted by a scaling process, in some way, to the distribution of the data in $\mathcal{H}$ and to adapt the width to the data spread area in $\mathcal{H}$: in areas with dense data distribution, the width should be small, and in sparse areas the width should be large.

By studying density vs. sparsity in $\mathcal{H}$ [8] argues that if a pattern x is in a dense area its neighbors are close to it, and conversly in sparse area. Thus, by considering the x's k-NNs (k-nearest neighbors), we get an index β of density of x's neighborhood. This index is

$$\beta(x) = (1/k) \sum_i k(x, x_i), x_i \in x\text{'s } k\text{-NNs} \qquad (8)$$

A large β means a dense area where x lies and vice versa, since in $\mathcal{H}$, the expression of the Euclidean distance of patterns x and y given by $d^2 = \langle \phi(x) - \phi(y), \phi(x) - \phi(y) \rangle = \|\phi(x) - \phi(y)\|^2 = \|\phi(x)\|^2 + \|\phi(y)\|^2 - 2\langle \phi(x), \phi(y) \rangle = 2(1 - k(x, y))$ shows that the value of $k(x, y)$ is inverse to d.

Selecting the x's k-NNs means that we are selecting the k patterns which have a distance d less than those of the remaining part in the training set, and this means that they have the highest kernel value $k(x, .)$. In the following, the Gaussian RBF kernel is formulated as $k(x, y) = \exp\big[-\lambda(x, y)\|x-y\|^2/2\sigma^2\big]$, [8]. Each pattern x has an individual value $\tilde{\lambda}(x)$. If $\beta(x)$ is large, i.e. in dense area, a large $\tilde{\lambda}(x)$ will increase $\lambda(x, y)$ which means decreasing the width σ. Conversely, a small $\beta(x)$ means that a small $\tilde{\lambda}(x)$ is used. Thus, the modified Gaussian RBF kernel is formulated as

$$k(x, y) = \exp(-\tilde{\lambda}(x) \cdot \tilde{\lambda}(y) \cdot \lambda \|x - y\|^2 / 2\sigma^2) \qquad (9)$$

In this way, $\lambda(x, y)$ fluctuates within a very small range around 1 to reduce distorting the metric induced by the RBF kernel.

In order to formulate the modified Gaussian RBF kernel as indicated in Eq. (9), we need first to compute $\tilde{\lambda}(\cdot)$.

Algorithm 2: Scaling Kernel width

For each training set $\mathcal{X}_i, i = 1, \ldots, N$, do:

- For each training pattern $x_{i,j} \in \mathcal{X}_i$, $j = 1, \ldots, m_i$, the following steps are needed:
  - Find its k-NNs in $\mathcal{X}_i$, and compute $\beta(x_{i,j})$, according to Eq. (8)
  - Compute the mean of $\beta(x_{i,j})$

$$\breve{\beta} = (1/m_i) \sum_i \beta(x_{i,j}) \qquad (10)$$

  - Compute $\tilde{\lambda}(x_{i,j})$ by

$$\tilde{\lambda}(x_{i,j}) = 1 + \delta(\beta(x_{i,j}) - \breve{\beta}) \qquad (11)$$

    where $\delta \in [0, 1]$ is a factor for weighting intensity.

- For each testing pattern $z_i \in \mathcal{X}'$, $i = 1, \ldots, m'$
  - Find its k-NNs in $\mathcal{X}$, and compute $\beta(z_i)$, according to Eq. (8)
  - Compute $\breve{\beta}$ using Eq. (10).
  - Compute $\tilde{\lambda}(z_i)$, using Eq. (11).

For each pair (x,y), the modified Gaussian RBF kernel is formulated according to Eq. (9).

An important remark is that the dissimilarity in Eq. (7) is not properly defined anymore. However, if we understand the role $\lambda(x, y)$ as locally stretching the hypersphere surface in $\mathcal{H}$, the relative positions of the data in $\mathcal{H}$ is not deeply modified, and the interpretation of $d(\mathcal{X}, x)$ remains valid.

4. Experimental evaluation

For experimental evaluation, considering several sound libraries was necessary for building a representative, large, and sufficiently diversified database. The major part of the sound samples used in the recognition experiments was taken from different sound libraries [18, 24]. Some particular classes of sounds have been built or completed with sounds recorded by us. The selected sounds are impulsive and they are typical of surveillance applications. More precisely, the database consists of above 1,000 sounds belonging to 9 classes (Human screams, Gunshots, Glass breaks, Explosions, Door slams, Dog barks, Phone rings, ..., etc) as described in the following Table 1.

Table 1: Classes of sounds and number of samples in the database used for performance evaluation.

Classes	Train	Test	Total
Human screams	48	25	73
Gunshots	150	75	225
Glass breaks	58	30	88
Explosions	41	21	62
Door slams	209	105	314
Dog barks	36	19	55
Phone rings	34	17	51
Children voices	58	29	87
Machines	40	20	60
Total	674	341	1015

All signals in the database have a 16 bits resolution and are sampled at 44100 Hz. A feature vector (of dimension 12) issued from Mel Frequency Cepstral Coefficients (MFCCs) is computed for each entire sample file. The analysis window length for all features was 25 ms and the used windowing function was Hamming. The overlap between successive frames was 50% of the frame length.

The objective of this experimental study is to evaluate the quality of the approache described above. The sound recognition ability of the our method is evaluated by supervised classification using a 1-SVM-based classifier as described in Algorithm 1 in which the ν value is fixed to 0.2. Results for Algorithm 2 are displayed in Table 2. The weighting intensity factor δ is set to 0.5 in all experiments.

The results listed in Table 2 show that the performance with scaled width kernel is better than that with fixed width kernel and thus confirms that the pattern spatial distribution affects the sound recognition quality.

5. Conclusions

In this paper, we propose a system motivated by a practical surveillance application that uses a discriminative method based on one-class SVMs, together with a sophisticated dissimilarity measure. In order to enhance the discrimination ability of the proposed system, the kernel parameters are also studied. The under-fitting and over-fitting phenomenon is tackled by taking into account the patterns' spatial distribution.

References

[1] S. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions" *Neural Networks*, vol. 12, pages 783–789, 1999.

[2] S. Amari and S. Wu, "Conformal transformation of kernel functions: A data-dependent way to improve the performance of support vector machine classifiers" *Neural Processing Letter*, vol. 15, 2002.

[3] L. Breiman and S. Wu, "Bagging predictors", *Machine Learning*, vol. 24, pp 123–140, 1996.

[4] C. M. Bishop, "Novelty detection and neural networks validation," in *IEE Proceedings on Vision, Image and Signal Processing. Special Issue on Applications of Neural Networks*, May 1994.

[5] R. Burbidge, *Adaptive kernels for support vector classification*. Lecture Notes in Statistics, Issu. 171, pp. 345–356, Springer Verlag, NY, 2003.

[6] C. Campbell, P. Bennett, "A linear programming approach to novelty detection," in *NIPS 2000*, pages 395-401, 2000.

[7] O. Chapelle, V. Vapnik, O. Bousquet and S. Mukherjee, "Choosing multiple parameters for support vector machines", *Machine Learning*, vol. 46, pp. 131–160, 2002.

[8] Q. Chang, Q. Chen and X. Wang, "Scaling Gaussian RBF kernel width to improve SVM classification, " in *IEEE International conference on Neural Networks and Brain, (ICNNB'05)*, Oct 2005.

Table 2: *Recognition accuracy using fixed and scaled-width kernel applied to 1-SVMs based classifier.*

σ	acc. with fixed width	acc. with scaled-width
12	92.33	91.56
10	92.80	93.66
8	92.80	93.73
6	91.64	92.99
4	91.83	91.89
2	90.79	91.22
1	88.23	88.23
0.5	87.66	87.66
0.1	88.90	89.33

[9] F. Desobry, M. Davy & C. Doncarli, "An online kernel change detection algorithm," in *IEEE Trans. on Signal Processing*, August 2005, vol. 53, no. 8.

[10] M. Davy and S. Godsill, "Detection of Abrupt Spectral Changes using Support Vector Machines. An application to audio signal segmentation," in *IEEE ICASSP 2002*, Orlando, USA, May 2002.

[11] M. Davy and F. Desobry and S. Canu, "Estimation of Minimum Measure Sets in Reproducing Kernel Hilbert Spaces and Applications," in *IEEE ICASSP 2006*, Toulouse, France, May 2006.

[12] M. Davy, F. Desobry, A. Gretton and C. Doncarli, "An online Support Vector Machine for Abnormal Events Detection" *Journal of Signal Processing*, vol. 86, no. 8, pages 2009–2025, aug. 2006.

[13] A. Ganapathiraju, J. Hamaker and J. Picone, "Support Vector Machines for Speech Recognition," in *International Conference on Spoken Language Processing (ICSLP)*, Sydney, Australia, nov. 1998.

[14] N. Japkowicz, *Concept-Learning in the absence of counter-examples: an autoassociation-based approach to classification*. PhD. thesis, The State University of New Jersey, 1999.

[15] J. Kandola and J. Shawe-Taylor, "Refining kernels for regression and uneven classification problems," in *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, 2003.

[16] G. Karakoulas and J. Shawe-Taylor, "Optimizing classifiers for imbalanced training sets," in *Advanced in Neural Information Processing Systems*, 1999.

[17] S. S. Keerthi and C. -J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel", *Neural Computation*, 15(7), pp. 1667-1689, 2003.

[18] Leonardo Software. Santa Monica, USA. *http://www.leonardosoft.com.*

[19] Y. Lin ,Y. Lee and G. Wahba, "Support vector machines for classification in nonstandard situations", *Machine Learning*, vol. 46, pp 191–202, 2002.

[20] L. Manevitz, M. Youssef, "One-class SVMs for document classification," *Journal of Machine Learning Research*, vol. 2, 2001.

[21] M. Moya, M. Koch & al, "One-class classifier networks for target recognition applications," in *world congress on neural networks*, Portland, apr. 1993.

[22] A. Rabaoui, M. Davy, S. Rossignol, Z. Lachiri, N. Ellouze, "Using One-Class SVMs and Wavelets for Audio Surveillance Systems," *IEEE Transactions on information forensics and security*, submitted, Fev. 2007.

[23] G. Ratsch, T. Onoda and K. R. Miuller, "Soft margins for AdaBoost", *Machine Learning*, vol. 42, pp. 287–320, 2001.

[24] Real World Computing Paternship. CD-Sound scene database in real acoustical environments. *http://tosa.mri.co.jp/sounddb/indexe.html*

[25] G. Ritter, M. T. Gallegos, "Outliers in statistical pattern recognition and an application to automatic chromosome classification," in *Pattern Recognition Letters*, April 1997, pages 525-539.

[26] N. Smith and M. Gales, *Speech Recognition using SVMs*. T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, Advances in Neural Information Processing Systems 14. MIT Press, pages 1–8, 2002.

[27] B. Scholkopf, A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, 2002.

[28] D. M. J. Tax, *One-class classification*. PhD. thesis, Delft University of Technology, June 2001.

[29] Y. Li, H. Zaragoza, R. Herbrich, J. Shawe-Taylor, J. Kandola, "The perceptron algorithm with uneven margins," in *Proceedings of the Nineteeth International Conference on Machine Learning*, pages 379–386, 2002.