A NEW SUPPORT VECTOR MACHINE METHOD FOR MEDICAL IMAGE CLASSIFICATION

Trung Le, Dat Tran, Wanli Ma and Dharmendra Sharma

Faculty of Information Sciences and Engineering
University of Canberra, Australia
Email: {trung.le, dat.tran, wanli.ma, dharmendra.sharma}@canberra.edu.au

ABSTRACT

One of the important problems in medical imaging is two-class classification, for example determination of benign from malignant cases in breast cancer treatment. In this paper we present a new support vector machine method for two-class medical image classification. The key idea of this method is to construct an optimal hypersphere such that both the interior margin between the surface of this sphere and the normal data, and the exterior margin between this surface and the abnormal data are as large as possible. The proposed method is easily implemented and can reduce both false positive and false negative error rates to obtain very good classification results. Experiments were performed on three medical image data sets to evaluate the proposed method.

Index Terms— Pattern classification, Medical image processing, support vector machine.

1. INTRODUCTION

Medical imaging is a vital component of most of applications in the clinical track of events including clinical diagnostis settings, planning, consummation, and evaluation of surgical and radiotherapeutical procedures. Two global categories in imaging are anatomical and functional. Anatomical modalities include X-ray, CT (computed tomographya), MRI (magnetic resonance imagingb), US (ultrasoundc), portal images, and (video) sequences. Some prominent derivative techniques are so detached under a separate name, for example MRA (magnetic resonance angiography), DSA (digital subtraction angiography, derived from Xray), CTA (computed tomography angiography), and Doppler (derived from US, referring to the Doppler effect measured). Functional modalities depicting primarily information on the metabolism of the underlying anatomy, include (planar) scintigraphy, SPECT (single photon emission computed tomographyd), PET (positron emission tomographye), which together make up the nuclear medicine imaging modalities, fMRI (functional MRI), EEG (electro encephalography), and MEG (magneto encephalography) [20].

Consider medical imaging for breast cancer. Currently there are no methods to prevent breast cancer, therefore early detection represents a very important factor in cancer treatment to obtain a high survival rate. The most reliable method in early detection is mammography. However reading digital mammogram images is very difficult due to their low contrast and differences in the types of tissues [1]. Preliminary signs of masses and calcification clusters are important visual clues but they are very subtle and varied in appearance in the early stages of breast cancer. This is the main reason for the development of pattern classification systems to assist specialists in medical institutions. Recently, support vector machine (SVM) has emerged as a powerful pattern classification method for image classification [8]. SVM provides great results in the classification of high-dimensional datasets and pixel-based image classifiers [13, 2, 9, 10].

There are different SVM methods proposed for classification. Traditional SVM constructs an optimal separating hyperplane between classes by focusing on the training vectors close to the edge of the class descriptors. These training vectors are called support vectors. Other training vectors are discarded, therefore high classification accuracy can be achieved with small training data sets. In one-class classification also called novelty detection, one-class SVM (OCSVM) is used to capture the characteristic of normal data to construct a data description and then applies this description to detect abnormal data or outliers that cannot fit this description very well [21]. OCSVM constructs a hyperplane to separate the normal data such that the margin between the hyperplane and outliers is maximized [23] [24]. Recently, a small sphere and large margin (SSLM) approach has been introduced [32] to surround the normal data in an optimal hypersphere such that the margin—distance from outliers to the optimal hypersphere, is maximized. This SSLM approach is easily implemented, is helpful for parameter selection and provides very good detection results on a number of real data sets.

However the SSLM method can result in a very closed and tight boundary around the normal data and as a result, some normal data points can be outside the hypersphere resulting in a high false negative error rate. In order to over-

come this problem, we propose to have an additional margin between the surface of this hypersphere and the normal data. The key idea is to construct an optimal hypersphere such that both the interior margin between the surface of this hypersphere and the normal data and the exterior margin between this surface and the abnormal data are as large as possible. The proposed method is easily implemented and can reduce both false positive and false negative error rates. It can also be proved that the SSLM approach is a special case of the proposed approach. Experiments were performed on three medical image data sets for evaluation. In all of these experiments, high classification rates are achieved for the proposed method.

2. CURRENT SVM METHODS

2.1. One-Class Support Vector Machine (OCSVM)

In OCSVM [23], a hyperplane is determined to separate all normal data and at the same time maximise the margin between the normal data and the hyperplane. OCSVM can be modelled as follows

$$\min_{w,\rho} \Big(\frac{1}{2}||w||^2 - \rho + \frac{1}{\nu s}\sum_{i=1}^{s}\xi_i\Big) \tag{1}$$

subject to

$$\begin{aligned} &w^T\phi(x_i) \geq \rho - \xi_i \qquad i = 1,\ldots,s \\ &\xi_i \geq 0, \quad i = 1,\ldots,s \end{aligned} \tag{2}$$

where w is the normal vector of the hyperplane, ρ is the margin, ν is a positive constant, x_i, $i = 1,\ldots,s$ are data points, ξ_i, $i = 1,\ldots,s$ are slack variables, and $\phi(.)$ is a kernel function.

The decision function is $f(x) = sign(w^T\phi(x) - \rho)$. The unknown x is a normal data point if $f(x) = +1$ or an abnormal data point if $f(x) = -1$.

2.2. SVM Classification (SVMC)

SVMC was originally proposed to deal with the balanced data sets [24]. However, by selecting two appropriately proportional trade-off parameters, it can be used to deal with imbalanced datasets. Let x_i, $i = 1,\ldots,m_1$ be normal data points with label $y_i = +1$ and x_i, $i = m_1+1,\ldots,s$ be abnormal data points with label $y_i = -1$, and $m_2 = s - m_1$. SVMC can be modelled as follows

$$\min_{w,b} \Big(\frac{1}{2}||w||^2 + C_1\sum_{i=1}^{m_1}\xi_i + C_2\sum_{i=m_1+1}^{s}\xi_i\Big) \tag{3}$$

subject to

$$\begin{aligned} &y_i[w^T\phi(x_i) + b] \geq 1 - \xi_i \qquad i = 1,\ldots,s \\ &\xi_i \geq 0, \quad i = 1,\ldots,s \end{aligned} \tag{4}$$

where C_1, C_2 and b are real numbers. The decision function is $f(x) = sign(w^T\phi(x) + b)$.

2.3. Support Vector Data Description (SVDD)

SVDD [24] aims at drawing an optimal hypersphere containing normal data. Abnormal data are outside this hypersphere. The optimisation problem is as follows

$$\min_{R,\xi} \Big(R^2 + C_1\sum_{i=1}^{m_1}\xi_i + C_2\sum_{i=m_1+1}^{s}\xi_i\Big) \tag{5}$$

subject to

$$\begin{aligned} &||\phi(x_i) - c||^2 \leq R^2 + \xi_i \qquad i = 1,\ldots,m_1 \\ &||\phi(x_i) - c||^2 \geq R^2 - \xi_i \qquad i = m_1+1,\ldots,s \\ &\xi_i \geq 0, \quad i = 1,\ldots,s \end{aligned} \tag{6}$$

where R and c are radius and centre of the hypersphere, respectively. The decision function is $f(x) = sign(R^2 - ||\phi(x) - c||^2)$.

2.4. Small Sphere & Large Margin (SSLM)

The SSLM approach combines the ideas of OCSVM and conventional two-class SVM [29] in minimising a hypersphere containing all normal data and simultaneously maximising the margin which is the distance from outliers (abnormal data) to the surface of the optimal hypersphere. This SSLM approach can be formulated by the following optimisation problem:

$$\min_{R,c,\xi,\rho} \Big(R^2 - \nu\rho^2 + \frac{1}{\nu_1 m_1}\sum_{i=1}^{m_1}\xi_i + \frac{1}{\nu_2 m_2}\sum_{i=m_1+1}^{s}\xi_i\Big) \tag{7}$$

subject to

$$\begin{aligned} &||\phi(x_i) - c||^2 \leq R^2 + \xi_i \qquad i = 1,\ldots,m_1 \\ &||\phi(x_i) - c||^2 \geq R^2 + \rho^2 - \xi_i \qquad i = m_1+1,\ldots,s \\ &\xi_i \geq 0, \quad i = 1,\ldots,s \end{aligned} \tag{8}$$

where ν, ν_1 and ν_2 are three positive constants, ρ^2 is outside margin (distance from abnormal data to the surface of the hypershere).

It can be seen that minimising the cost function (7) will make the radius R as small as possible and the margin ρ^2 as large as possible. Therefore this approach is called Small Sphere and Large Margin (SSLM). The hypersphere only surrounds the positive class (normal data) and SSLM aims to find a large margin between this hypersphere and the abnormal data points. The decision function is $f(x) = sign(R^2 - ||\phi(x) - c||^2)$.

3. SS2LM: A NEW SVM METHOD

As mentioned above, the SSLM approach can produce a very closed and tight hypersphere boundary around the normal data. Therefore some normal data points can be outside the hypersphere and they will be classified as abnormal data resulting in a high false negative error rate.

To overcome this problem, we introduce an interior margin between the surface of the hypershere and normal data to the SSLM approach. This margin is proportional to the exterior margin between that surface and abnormal data by a proportionality constant δ. This constant is determined based on the ratio of normal data and abnormal data points. Our approach is to construct an optimal hypersphere such that both the interior and exterior margins are as large as possible.

The proposed method can be formulated by the following optimisation problem:

$$\min_{R,c,\xi,\rho} \Big(R^2 - \nu\rho^2 + \frac{1}{\nu_1 m_1}\sum_{i=1}^{m_1}\xi_i + \frac{1}{\nu_2 m_2}\sum_{i=m_1+1}^{s}\xi_i \Big) \quad (9)$$

subject to

$$\begin{array}{ll} ||\phi(x_i) - c||^2 \le R^2 - \delta\rho^2 + \xi_i & i = 1,\ldots,m_1 \\ ||\phi(x_i) - c||^2 \ge R^2 + \rho^2 - \xi_i & i = m_1+1,\ldots,s \\ \xi_i \ge 0, \quad i = 1,\ldots,s & \end{array} \quad (10)$$

It can be seen that minimising the cost function (9) will make the radius R as small as possible and the margin ρ^2 as large as possible, resulting in the proposed margin $\delta\rho^2$ is also as large as possible.

The following Lagrange is introduced to investigate the proposed problem

$$\begin{aligned} L(R,c,\xi,\alpha,\beta) &= R^2 - \nu\rho^2 \\ &+ \frac{1}{\nu_1 m_1}\sum_{i=1}^{m_1}\xi_i + \frac{1}{\nu_2 m_2}\sum_{i=m_1+1}^{s}\xi_i \\ &+ \sum_{j=1}^{s}\alpha_j(y_j||\phi(x_j) - c||^2 - y_j R^2 - z_j\rho^2 - \xi_j) \\ &- \sum_{k=1}^{s}\beta_k\xi_k \end{aligned} \quad (11)$$

where $y_i = +1, i = 1,\ldots,m_1$, $y_i = -1, i = m_1+1,\ldots,s$, and $z_i = \frac{1}{2}[(1-y_i) + (1+y_i)\delta], i = 1,\ldots,s$.
where $\alpha_i \ge 0$ and $\beta_i \ge 0$ are Lagrange multipliers. Setting derivatives of $L(R,c,\xi,\alpha,\beta)$ with respect to primal variables to 0, we obtain the dual form

$$\min_{\alpha} \Big(\sum_{i=1}^{s}\alpha_i\alpha_j y_i y_j K(x_i,x_j) - \sum_{i=1}^{s}\alpha_i y_i K(x_i,x_i) \Big) \quad (12)$$

subject to

$$\begin{array}{l} 0 \le \alpha_i \le \dfrac{1}{\nu_1 m_1}, \quad i = 1,\ldots,m_1 \\ 0 \le \alpha_j \le \dfrac{1}{\nu_2 m_2}, \quad j = m_1+1,\ldots,s \\ \displaystyle\sum_{i=1}^{s}\alpha_i y_i = 1, \qquad \sum_{i=1}^{s}\alpha_i = \frac{2\nu + 1 - \delta}{\delta + 1} \end{array} \quad (13)$$

The dual form is also a quadratic optimization problem and has the same form as the dual of the ν-SVM [22], thus it can be solved with the ν-SVM solver in the LIBSVM software [6].

To classify an unknown data point x, the following decision function is used

$$\begin{aligned} f(x) = sgn\Big(R^2 - \sum_{i=1}^{s}\sum_{j=1}^{s}\alpha_i\alpha_j y_i y_j K(x_i,x_j) - K(x,x) \\ + 2\sum_{k=1}^{s}\alpha_k y_k K(x,x_k) \Big) \end{aligned} \quad (14)$$

x is a normal data point if $f(x) = +1$ or an abnormal data point if $f(x) = -1$

The parameters R, c and ρ are calculated as follows

$$\begin{aligned} R^2 &= \frac{1}{n_1}P_1 \\ ||c||^2 &= \sum_{i=1}^{s}\sum_{j=1}^{s}\alpha_i\alpha_j y_i y_j K(x_i,x_j) \\ \rho^2 &= \frac{1}{n_2}P_2 - \frac{1}{n_1}P_1, \end{aligned} \quad (15)$$

where

$$\begin{array}{c} n_1 = |S_1|, \ S_1 = \Big\{ x_i \mid 0 < \alpha_i < \dfrac{1}{\nu_1 m_1}, \ 1 \le i \le m_1 \Big\} \\ n_2 = |S_2|, \ S_2 = \Big\{ x_j \mid 0 < \alpha_j < \dfrac{1}{\nu_2 m_2}, \ m_1+1 \le j \le s \Big\} \\ P_1 = \displaystyle\sum_{x_i \in S_1}\Big[K(x_i,x_i) + ||c||^2 - 2\sum_{k=1}^{s} y_k\alpha_k K(x_k,x_i) \Big] \\ P_2 = \displaystyle\sum_{x_i \in S_2}\Big[K(x_i,x_i) + ||c||^2 - 2\sum_{k=1}^{s} y_k\alpha_k K(x_k,x_i) \Big] \end{array} \quad (16)$$

4. ν-PROPERTY AND MARGIN ERRORS

Similar to SSLM and ν-SVM, a training data point x_i is called a Support Vector (SV) if the corresponding $\alpha_i > 0$, and it is called a Margin Error (ME) if the corresponding slack variable $\xi_i > 0$.

Let m^+ and m^- denote the number of MEs in the normal and abnormal data, s^+ and s^- denote the number of SVs in

the normal and abnormal data, respectively. Then, for parameters ν, ν_1, and ν_2, we have

$$\begin{aligned} \frac{m^+}{m_1} &\leq \frac{(\nu+1)\nu_1}{\delta+1} \leq \frac{s^+}{m_1} \\ \frac{m^-}{m_2} &\leq \frac{(\nu-\delta)\nu_2}{\delta+1} \leq \frac{s^-}{m_2} \end{aligned} \tag{17}$$

Proof: From (11), we have

$$\begin{aligned} \frac{\partial L}{\partial R} = 0 \quad &\Rightarrow \quad \sum_{i=1}^{s} \alpha_i y_i = 1 \\ \frac{\partial L}{\partial \rho} = 0 \quad &\Rightarrow \quad \delta \sum_{i=1}^{m_1} \alpha_i + \sum_{i=m_1+1}^{s} \alpha_i = \nu \end{aligned} \tag{18}$$

Summing up α_i over the positives MEs leads to

$$\frac{m^+}{\nu_1 m_1} \leq \sum_{i=1}^{m_1} \alpha_i = \frac{\nu+1}{\delta+1} \leq \frac{s^+}{m_1 \nu_1} \tag{19}$$

The second one of (17) can be proven in a similar manner.

It can be seen that when δ increases, both the margin errors m^+ and m^- decrease. When $\delta = 0$, we obtain the margin errors for the SSLM approach [32].

5. EXPERIMENTAL RESULTS

5.1. Breast Cancer Data Set

The Wisconsin breast cancer data set consists of 683 feature vectors of which 444 vectors are labeled $Benign$ and 239 vectors were $Malignant$. Determination of benign from malignant cases is an important problem in breast cancer treatment. Each vector has 9 features including clump thickness (the extent to which epithelial cell aggregates were mono- or multi-layered), uniformity of cell size, uniformity of cell shape, marginal adhesion (cohesion of the peripheral cells of the epithelial cell aggregates), single epithelial cell size, bare nuclei (the proportion of single epithelial nuclei that were devoid of surrounding cytoplasm), blandness of nuclear chromatin, normal nucleoli, and mitoses. All of these feature values are continuous and range from 1 to 10. A cancer image for digital analysis was generated by a video camera mounted on top a microscope. A slide was projected into the camera and the image was captured. The resulting image is stored in memory as a 2-dimensional array, with each pixel having a value between 0 and 255 representing the light intensity at that point [31].

5.2. Liver Disorders Data Set

There are many disorders of the liver that require clinical care by a physician or other healthcare professional, for example alcohol-induced liver disease, chronic liver disease, congenital defects, and hepatitis. The BUPA liver disorders data set consists of 345 feature vectors of multiclasses. Each vector has 7 features including mean corpuscular volume, alkaline phosphotase, alamine aminotransferase, aspartate aminotransferase, gamma-glutamyl transpeptidase, number of half-pint equivalents of alcoholic beverages drunk per day, and field used to split data into two sets. The first 5 features are all blood tests which are sensitive to liver disorders that might arise from excessive alcohol consumption [27]. In this paper, we used a version of this data set downloadable at [28] containing only 6 continuous features.

5.3. SPECTF Heart Data Set

Single Proton Emission Computed Tomography (SPECT) imaging is used as a diagnostic tool for myocardial perfusion. The patient is injected with radioactive tracer then a stress image was taken 10-15 minutes after injection during maximal stress, and a rest image 2-5 hours after injection. Cardiologists compare stress and rest studies in order to detect abnormalities in the left ventricle perfusion. Normally the SPECT images are presented to a cardiologist as three sets of two-dimensional images [15].

The SPECTF heart data set [25] describes diagnosing of cardiac SPECT images. There are 349 image sets (patients) in this data set. Each of the patients is classified into two categories: normal and abnormal. A total of 44 continuous features was created for each patient . In our experiment, 254 abnormal patients were used as target class and 95 normal patients as outlier class.

5.4. Experiments

We performed classification experiments on these 3 data sets to compare OCSVM, SVM-C, SVDD, SSLM, and SS2LM methods. These sets are balanced so we created at random imbalanced subsets such that the ratio of number of normal data points (m_1) and number of abnormal data points (m_2) was 19:1, i.e. 95% of data points are normal and 5% are abnormal. Creating subsets were repeated 10 times and the average classification rate for 10 times was determined.

Table 1. Number of data points in the 3 data sets. #pos: total positive data points, #neg: total negative data points, m_1: positive data points for training, m_2: negative data points for training, and d: dimension. The remaining data points $(\#pos - m_1)$ and $(\#neg - m_2)$ were used for testing.

Data set	#pos	#neg	m_1	m_2	d
Breast Cancer	444	239	355	18	10
Liver Disorders	200	145	116	6	6
Spectf Heart	254	95	203	10	44

The classification rate acc is measured as [14]

$$acc = \sqrt{acc^{+}acc^{-}} \tag{20}$$

where acc^{+} and acc^{-} are the classification accuracy on normal and abnormal data, respectively.

The popular RBF kernel function $K(x, x') = e^{-||x-x'||^2/\gamma}$ was used in our experiments to compare with the SSLM approach. We used the same parameter settings suggested in [32], the parameter γ was searched in $\{\sigma_0^2/16, \sigma_0^2/8, \sigma_0^2/4, \sigma_0^2/2, \sigma_0^2, 2\sigma_0^2, 4\sigma_0^2, 8\sigma_0^2, 16\sigma_0^2\}$, where σ_0^2 is the mean norm of the training data. For SVMs the penalty parameters C, C_1 and C_2 were searched over the grid $\{0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100, 500\}$, such that the ratio C_2/C_1 belonged to

$$\left\{\frac{1}{4} \times \frac{m_1}{m_2}, \frac{1}{2} \times \frac{m_1}{m_2}, \frac{m_1}{m_2}, 2 \times \frac{m_1}{m_2}, 4 \times \frac{m_1}{m_2}\right\} \tag{21}$$

For OCSVM, the parameter ν was searched in $\{0.01k, 0.1k\}$, where k was an integer ranging from 1 to 9. For SSLM and SS2LM, the parameter ν was searched in $\{10, 30, 50, 70, 90\}$, while ν_1 and ν_2 were selected from $\{0.001, 0.01\}$.

Table 2. Classification results (in %) on the 3 data sets

Method	Breast Cancer	Liver Disorders	SPECTF Heart
OCSVM	95.0	73.8	84.1
SVM-C	88.9	81.7	88.4
SVDD	89.9	78.6	89.8
SSLM	98.7	85.1	90.6
SS2LM	**99.0**	**87.4**	**91.4**

We can see from Table 2 that the SSLM method provides better performance than the OCSVM, SVM-C and SVDD methods. The proposed SS2LM approach achieved higher classification rates than the SSLM method for all of the datasets because it produces lower false negative error rates.

6. CONCLUSION

We have analysed current SVM methods for medical image classification and propose a new SVM method to provide a lower false negative error rates in two-class medical image classification. We have evaluated the proposed method on three popular data sets which were breast cancer, liver disorders and SPECTF heart images. The experimental results showed a good performance for the proposed method. For further investigation, larger real data sets will be used for evaluation. A fuzzy approach to this SVM method will be investigated to reduce the sensitivity of SVM to noisy data and outliers.

7. REFERENCES

[1] M.-L. Antonie, O.R. Zaiane, and A. Coman, "Application of Data Mining Techniques for Medical Image Classification", Proceedings of MDM/KDD, pp. 94-101, 2001.

[2] M. Brown, H.G. Lewis, and S.R. Gunn, "Linear Spectral Mixture Models and Support Vector Machines for Remote Sensing," IEEE Transactions On Geoscience And Remote Sensing, vol. 38, no. 5, 2000

[3] C.M. Bishop, "Novelty detection and neural network validation," IEEE Proceedings-Vision, Image, and Signal Processing, 1994

[4] V. Barnett and T. Lewis, *Outliers in statistical data* 3rd ed. Wiley, 1978

[5] C. Campbell & K.P. Bennet, "A linear programming approach to novelty detection," *Advances in neural information processing*, 2001

[6] C.-C. Chang and C.-J. Lin, LIBSVM: A Library for Support Vector Machines. The LIBSVM software package is available at http://www.csie.ntu.edu.tw/ cjlinlibsvm

[7] C. Chuang, S. J. Su Jeng, and C. Hsiao, "Robust support vector regression networks for function approximation with outliers," *IEEE Trans. Neural Netwroks*, *13*, 13221330, 2002

[8] Y. Fan, D. Shen and C. Davatzikos "Classification of Structural Images via High-Dimensional Image Warping, Robust Feature Extraction, and SVM," Proceedings of MACCAI, pp. 1-8, 2005

[9] G.M. Foody and A. Mathur, "A Relative Evaluation of Multiclass Image Classification by Support Vector Machines," IEEE Transactions On Geoscience And Remote Sensing, vol.42, no. 6, 2004

[10] J.A. Gualtieri and R. F. Cromp, "Support vector machines for hyperspectral remote sensing classification," Proceedings of the SPIE, vol. 3584, pp. 221232, 1999

[11] C.-W. Hsu and C.-J. Lin "A comparison of methods for multiclass support vector machines," IEEE Trans. Neural Networks, vol. 13, pp. 415-425, 2002

[12] C.-C. Hsu, M.-F. Han, S.-H. Chang, H.-Y. Chung, "Fuzzy Support Vector Machines with the Uncertainty of Parameter C," Expert Systems With Application, vol. 44, pp. 6654-6658, 2009

[13] C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessement of support vector machines for land cover classifiocation," Int. J. Remote sensing, vol. 23, no. 4, pp. 725-749, 2002

[14] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training set: One-sided selection," Proc. 14th International Conference on Machine Learning, 1997

[15] L.A. Kurgan, K.J. Cios, R. Tadeusiewicz, M. Ogiela and L.S. Goodenday, "Knowledge Discovery Approach to Automated Cardiac SPECT Diagnosis," Artificial Intelligence in Medicine, vol. 23:2, pp. 149-169, 2001

[16] Z. Lao, D. Shen, Z. Xue, B. Karacali, S.M. Resnick, C. Davatzikos, "Morphological classification of brains via high-dimensional shape transformations and machine learning methods," Neuro Image vol. 21, pp. 46-57, 2004

[17] G.H. Lee, J. S. Taur and C.W. Tao, "A Robust Fuzzy Support Vector Machine for Two-class Pattern Classification," International Journal of Fuzzy Systems, vol. 8, no. 2, pp. 76-87, 2006

[18] C.-F. Lin and S.-D. Wang, "Fuzzy Support Vector Machines," IEEE Transactions on Neural Networks, vol. 13, no. 2, 464-472, 2002

[19] Y. Lin, Y. Lee and G. Wahba, "Support vector machine for classification in nonstandard situations," Machine Learning, vol.15, pp. 1115-1148, 2002

[20] J. B. A. Maintz and M. A. Viergever, "A Survey of Medical Image Registration," Medical Image Analysis, Oxford University Press, vol. 2, no. 1, pp 1-37, 1998

[21] M.M. Moya, M.W. Koch and L.D. Hostetler, "One-class classifier networks for target recognition applications," Proceedings of world congress on neural networks, pp. 797-801, 1991

[22] S. Roberts and L. Tarassenko, "A Probabilistic Resource Allocation Network for Novelty Detection," Neural Computation, vol. 6, pp.270-284, 1994

[23] Schlkopf and A. J. Smola, *Learning with kernels* Cambridge: The MIT Press, 2002

[24] D. M. J. Tax and R. P. W. Duin, "Support vector data description," Machine Learning, vol. 54, pp. 45-56, 2004

[25] D. M. J. Tax, Data sets downloadable at http://ict.ewi.tudelft.nl/ davidt/occ/index.html

[26] G.G. Towel, "Local expert autoassociator for anomaly detection," Proc. 17th International Conference on Machine Learning, 2000

[27] http://archive.ics.uci.edu/ml/index.html

[28] http://www.csie.ntu.edu.tw/ cjlin/libsvmtools/datasets/

[29] V. Vapnik, "The nature of statistic learning theory," New York: Springer Verlag, 1995

[30] J. Vert and J.P. Vert, "Consistency and convergence rates of one class SVM and related algorithm," Journal of Machine Learning Research, vol. 7, pp. 817-854, 2006

[31] W.H. Wolberg, and O.L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," Proceedings of the National Academy of Sciences, vol. 87, pp. 9193–9196, 1990

[32] M. Yu and J. Ye, A Small Sphere and Large Margin Approach for Novelty Detection Using Training Data with Outliers IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 31, pp. 2088-2092, 2009