Image Database Retrieval with Support Vector Data Descriptions

David M.J. Tax[1], Robert P.W. Duin[1], Kieron Messer[2]

[1]Pattern Recognition Group, Delft University of Technology
Lorentzweg 1, 2628 CJ Delft, The Netherlands
[2]Centre for Vision, Speech and Signal Processing
University of Surrey Guildford, Surrey, GU2 5XH, United Kingdom.
davidt@ph.tn.tudelft.nl

Keywords: pattern recognition, one-class problems, outlier detection, Support Vector Machines, Support Vector Data Description

Abstract

In previous research the Support Vector Data Description is proposed to solve the problem of One-Class classification. In One-Class classification one set of data, called the target set, has to be distinguished from the rest of the feature space. This description should be constructed such that objects not originating from the target set, by definition the outlier class, are not accepted by the data description. In this paper the Support Vector Data Description is applied to the problem of image database retrieval. The user selects an example image region as target class and resembling images from a database should be retrieved. This application shows some of the weaknesses of the SVDD, particularly the dependence on the scaling of the features. By rescaling features and combining several descriptions on well scaled feature sets, performance can be significantly increased.

1 Introduction

When one class of the data is to be distinguished from the rest of the feature space, a closed boundary around this set should be defined. This One-Class classification is often solved using density estimation or a model based approach. In this paper we use a method inspired by the Support Vector Classifier [Vap98]. Instead of using a hyperplane to distinguish between two classes, a hypersphere around the target set is used. This method is called the Support Vector Data Description (SVDD)[TD98].

In general the problem of One-Class classification is harder than the problem of normal Two-Class classification. For normal classification the decision boundary is supported from both sides by examples of each of the classes. Because in the case of One-Class classification only one set of data is available, only one side of the boundary is covered. Only on the basis of one class it is hard to decide how strictly the boundary should fit around the data in each of the directions. Even harder to decide is what the optimal scaling of the features should be to find the best separation of the target and outlier class.

To investigate the impact of this problem in a real application, the SVDD will be applied to an image database retrieval problem in which the user defines some interesting and desired image regions, and the application should retrieve the resembling images from the database. Only the user indicated regions will be used in training. The database is given by [Mes99]. Several features will be available, both color and texture features. In literature it is well known that color features perform very well in distinguishing the target images from the rest of the database[AKJ98]. Some very advanced techniques for image database retrieval have been used to use both types of features. A well optimized retrieval procedure, including automatic feature selection and extraction is also found in [Mes99].

In this paper we focus on the problem of the scaling of the data. In section 2 we will explain the Support Vector Data Description. In section 3 some characteristics of the database and information about the queries will be given. In section 4 the results of the queries by the SVDD will be shown and we end with discussions in section 5.

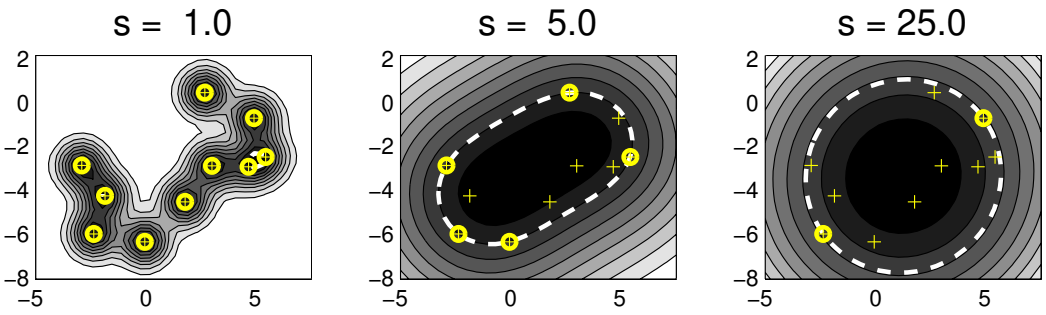

Figure 1: The Support Vector Data Description with a Gaussian kernel, applied to a simple 2 dimensional banana dataset.

2 Support Vector Data Description

For description of the domain of a dataset we capture it with a hypersphere with minimum volume. By minimizing the volume of the 'captured' feature space, we hope to minimize the chance of accepting outliers. Inspired by the Support Vector Method by Vapnik (see [Vap95], or for a more simple introduction [TdRD97]) one can extend this idea to determine an arbitrary shaped region in the original feature space, the Support Vector Data Description (SVDD) method [TD99].

Assume we a data set containing N data objects, $\{x_i, i = 1, .., N\}$ and the sphere is described by center a and radius R. To allow the possibility of outliers in the training set, the distance from $\mathbf{x}_i$ to the center $\mathbf{a}$ should not be strictly smaller than R^2, but larger distances should be penalized. Therefore we introduce slack variables ξ_i. We now minimize an error function containing the volume of the sphere and the distance from the boundary of the outlier objects. The constraints that objects are within the sphere are imposed by applying Lagrange multipliers:

$$L(R, a, \alpha_i) = R^2 + C \sum_i \xi_i - \sum_i \alpha_i \{R^2 - (x_i^2 - 2ax_i + a^2)\} - \sum_i \gamma_i \xi_i$$

with Lagrange multipliers $\alpha_i \geq 0$ and $\gamma_i \geq 0$. This function has to be minimized with respect to R, a and ξ_i and maximized with respect to α_i and γ_i.

Setting the partial derivatives of L to R and a to zero, gives:

$$\begin{aligned} \sum_i \alpha_i &= 1 \\ a &= \frac{\sum_i \alpha_i \mathbf{x}_i}{\sum_i \alpha_i} = \sum_i \alpha_i \mathbf{x}_i \\ \frac{\partial L}{\partial \xi_i} &= C - \alpha_i - \gamma_i = 0 \end{aligned} \quad (1)$$

From the last equation $\alpha_i = C - \gamma_i$ and because $\alpha_i \geq 0, \gamma_i \geq 0$, Lagrange multipliers γ_i can be removed when we demand that

$$0 \leq \alpha_i \leq C \quad (2)$$

Resubstituting these values in the Lagrangian gives to maximize with respect to α_i:

$$L = \sum_i \alpha_i (x_i \cdot x_i) - \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \quad (3)$$

with $0 \leq \alpha_i \leq C$, $\sum_i \alpha_i = 1$.

This function is a standard Quadratic Optimization problem and it should be maximized with respect to α_i. In practice it appears that a large fraction of the α_i becomes zero. For a small fraction $\alpha_i > 0$ and the corresponding objects are called Support Objects. We see that the center of the sphere depends just on the few support objects. The objects with $\alpha_i = 0$ can be disregarded.

Object z is accepted when:

$$\begin{aligned} (z - a)(z - a)^T &= (z - \sum_i \alpha_i x_i)(z - \sum_i \alpha_i x_i) \\ &= (z \cdot z) - 2 \sum_i \alpha_i (z \cdot x_i) + \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \\ &\leq R^2 \end{aligned} \quad (4)$$

In general this does not give a very tight description. The model of a hypersphere is assumed, which will not be satisfied in the general case. Analogous to the method of Vapnik [Vap98], we can replace the inner products $(x \cdot y)$ in equations (3) and in (4) by kernel functions $K(x, y)$ which gives a much more flexible method. When we replace the inner products by Gaussian kernels for instance, we obtain:

$$(x \cdot y) \rightarrow K(x, y) = \exp(-(x - y)^2 / s^2) \quad (5)$$

Equation (3) now changes into:

$$L = 1 - \sum_i {\alpha_i}^2 - \sum_{i \neq j} \alpha_i \alpha_j K(x_i, x_j) \quad (6)$$

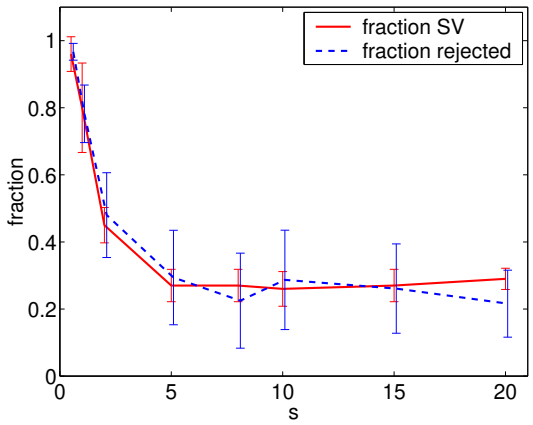

Figure 2: The number of Support Vectors and the error on the target set for a range of width parameters s.

and and the formula to check if a new object z is within the sphere (equation (4)) becomes:

$$1 - 2\sum_i \alpha_i K(z, x_i) + \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) \leq R^2 \tag{7}$$

In figure 1 the resulting data description is shown for a small artificial banana shaped dataset. The width parameter s ranges from $s = 1.0$ in the left subplot to $s = 25$ in the right subplot. For small values of s the SVDD resembles a Parzen density estimation, while for large s the original hypersphere solution is obtained[TD99]. Note also that the number of support vectors decreases.

So a more flexible description than the rigid sphere description is obtained. This Gaussian kernel contains one extra free parameter, the width parameter s in the kernel (equation (5)). As shown in [TD99] this parameter can be set by setting a priori the maximal allowed rejection rate of the target set, i.e. the error on the target set. This error can be estimated by the number of support vectors:

$$E[P(\text{error})] = \frac{\#SV}{N} \tag{8}$$

where $\#SV$ is the number of support vectors. The number of support vectors can be regulated by changing the width parameter s and therefore we can also set the error on the target set on the prespecified value. This is shown in figure 2 for the simple banana dataset. For changing s the data description is trained and the number of support vectors and the number of target objects (drawn independently from the target distribution) is measured. The fraction of support vectors closely follows the error on the target set.

The upper bound C on α_i can be set when an assumption on the fraction of outliers in the target set is made. Note we cannot set a priori restrictions on the error on the outlier class. In general we only have a good representation of the target class and the outlier class is per definition everything else.

An extra feature of the SVDD is the ability to include some example outliers in training. These negative examples are trained to be *outside* the data description and can make the data description tighter. To allow for errors of these example outliers (errors are outliers which are accepted by the description) also slack variables ξ_i and C have to be introduced. The relative sizes of the C for the target objects and the outlier objects then indicates the weighting of both errors by the SVDD. The proper choice of these variables C has to be made by the user.

3 Image database

The SVDD does not assume a certain distribution for the target class, but it requires well scaled data. That means that distances in feature space are more or less homogeneous. To investigate if this assumption is satisfied in practical situations, we will apply the method to the image database from Messer[Mes99]. The image database contains 3483 images, ranging from capture TV images to images from the MPEG-7 test set. The user chooses interesting regions in target image. On the selected image patches a One-Class classifier is trained and finally the resembling (both in color and texture) images should be retrieved from the database. In practical applications the speed of the retrieval is important, but in this paper we only will focus on the retrieval accuracy.

In figure 3 five test queries are defined (originally from [Mes99]). For each of the target images the user has defined 2 regions. To test the performance of the One-Class classifiers, also 5 to 8 other target images are defined, which resemble the target images. To give a ranking of the images, an error has to be defined. First the index should be accepted by the description. Further the distance to the center of the sphere, and the region size covered by the corresponding index is used to obtain a ranking for the accepted indices. The errors which are used, will be explained in section 4.

To accelerate the query, color and texture features of the image are calculated beforehand. For each pixel in each image 33 features are calculated. They are listed in table 1.

To further lower the computational burden in the processing of the complete image database,

Figure 3: The five defined queries, Cathedral, Queen, Hut, Flag, Newsreader, each with two user defined regions.

set	dim	type of feature	σ
1	9	discrete cosine transform	$1 \cdot 10^4$
2	8	gabor filters	$2 \cdot 10^3$
3	3	energy (green/int/red)	$3 \cdot 10^{-2}$
4	3	entropy (green/int/red)	$6 \cdot 10^{-2}$
5	3	mean (green/int/red)	$2 \cdot 10^0$
6	3	variance (green/int/red)	$6 \cdot 10^2$
7	4	wavelet transform	$1 \cdot 10^3$

Table 1: List of the 33 features, divided in 7 feature sets. Last column gives the average standard deviation of the data obtained from the brick wall region from the Cathedral image.

all images are segmented by clustering the set of pixels in the 33 dimensional feature space. On average about 30 to 50 cluster centers are obtained for each of the images. These cluster centers will be called indices. This reduces the images from about 300×200 pixels to about 50 indices. The drawback is that this only considers an average of the cluster and the variance structure of the cluster is lost. An example is given in figure 4.

4 Experiments

In all experiments the Support Vector Data Description is trained on the user defined regions such that about 10% of the training pixels is expected to fall outside the description (estimated by equation 8). The training time used by the Quadratic Optimization routine was restricted by

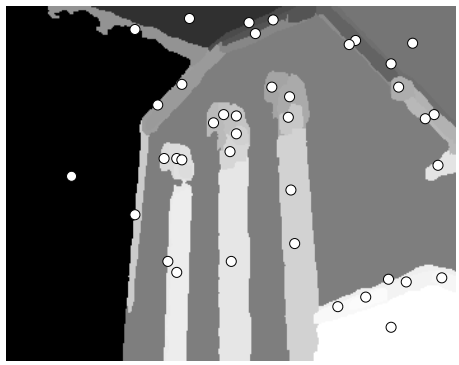

Figure 4: Clustered (or segmented) Cathedral image, containing 38 clusters. The cluster centers (indices) and they approximate positions are given by the dots.

using just 200 target objects (i.e. pixels randomly drawn from the user defined regions) per description. Using larger training sets gave sometimes somewhat better results, but required very long training times (longer than 10 minutes).

4.1 Normal features

First an SVDD is trained on the original 33 features, without preprocessing, to give an indication of how difficult the problem is. In table 3 the definitions of the errors is shown. First it is determined which indices are accepted by the SVDD. After that the accepted indices are ranked

		all features		separate features		
name query	error	no sc.	scaling	no sc.	scaling	best rankings
Cathedral	1	202 (181)	8 (7)	587 (21)	573 (17)	
	2	202 (106)	8 (7)	1978 (335)	1947 (254)	33 (28)
(5 target	3	202 (181)	8 (7)	429 (402)	430 (402)	
images)	4	202 (181)	8 (7)	106 (48)	35 (22)	
	5			107 (96)	37 (33)	
	6			107 (16)	37 (5)	
Queen	1	449 (141)	113 (31)	577 (509)	323 (259)	
	2	289 (158)	90 (29)	76 (59)	114 (91)	26 (14)
(5 target	3	306 (157)	150 (32)	705 (654)	611 (594)	
images)	4	354 (114)	129 (23)	153 (28)	111 (109)	
	5			153 (28)	111 (109)	
	6			153 (21)	111 (109)	
Hut	1	- (-)	- (-)	139 (97)	246 (243)	
	2	- (-)	- (-)	2164 (1697)	2121 (1553)	81 (49)
(5 target	3	- (-)	- (-)	195 (107)	194 (104)	
images)	4	- (-)	- (-)	770 (769)	764 (763)	
	5			770 (769)	764 (763)	
	6			770 (768)	764 (763)	
Flag	1	3215 (3214)	2602 (2601)	1165 (1110)	1153 (1101)	
	2	3215 (3214)	2602 (2601)	1536 (275)	1602 (331)	17 (7)
(8 target	3	3215 (3214)	2602 (2601)	3249 (2622)	3246 (2533)	
images)	4	3215 (3214)	2602 (2601)	344 (111)	369 (201)	
	5			597 (126)	711 (354)	
	6			1482 (1480)	1740 (1739)	
Newsreader	1	407 (287)	16 (9)	89 (72)	107 (68)	
	2	101 (94)	6 (5)	107 (102)	211 (203)	5 (4)
(5 target	3	202 (178)	8 (7)	2443 (1887)	2456 (2035)	
images)	4	382 (274)	13 (9)	5 (4)	6 (4)	
	5			20 (14)	23 (15)	
	6			7 (6)	26 (7)	

Table 2: Query results for the different data processing procedures: one SVDD using all features (normal and scaled to unit variance), descriptions for separate feature sets (also normal and scaled). The number indicates the position of the worst ranked desired image. The number in the brackets gives the second worst ranking. A dash indicates that the image is rejected by the description(s).The errors are defined in tables 3 and 4.

according to distance to center of the SVDD, the region size of the index or a combination of both. Finally these values are summed (of multiplied in error 4) over all indices to obtain a final score for the image.

nr	error
1	sum over region sizes of accepted indices
2	sum over distances from accepted indices to centers
3	sum over distance $\times$ (1+region size)
4	product over region sizes

Table 3: Errors defined for queries where all features are used in one SVDD

In the third column in table 2 the results are shown for the five query images on the complete image database. Because the database consist of almost 3500 images, a worst rank of 3215 means that the SVDD on that image is no better than random guessing. The results show that the four different errors (table 3) give about the same ranking of the images. Results on the Cathedral, Queen and Newsreader queries are acceptable. Although the desired images are not within the first 100, but the SVDD clearly performs better than random guessing. The results on the Hut and Flag queries on the other hand are disappointing. It appeared that in the Hut query only the training image is recognized well (it is ranked $91, 4, 5, 43$ in the four error measures). All other images are or

ranked very low or are completely rejected.

In the Flag query, most of the desired images are ranked very high. Only two images are not well recognized. These images are shown in figure 5. It appeared that these images are rotated over 90 degrees. This transformation is not considered irrelevant, and therefore the images are rejected.

Figure 5: Images with very poor ranking in the Flag query

The last column in table 1 shows the standard deviations of the data extracted from the wall region from the Cathedral image. Clearly the large difference in scale of the features can deteriorate the performance of the SVDD. Therefore the data is rescaled to unit variance in the target set and the experiments are repeated. The results are shown in the fourth column in table 2. The results on the Cathedral, Queen and Newsreader queries are very good. When the second worst rank is considered in the Queen query, all ranks are lower than 32. Hut and Flag suffer from the same problems as before. Even more so, the only image in the Hut query which is accepted by the SVDD is the original target image: *all* other images are rejected.

4.2 Separating the feature sets

To improve performance, especially in the Flag query, separate Support Vector Data Descriptions are trained on the individual feature sets (see table 1). It is hoped that it will not only make training an SVDD easier and more efficient, it will also give the possibility to use several feature combination rules. In the Flag query this means that the One-Class classifier can be forced to use the texture information.

	distance to center	region size	weighted by train perf.	feature sets	indices
1	X			sum	sum
2		X		sum	sum
3			X	sum	sum
4	X			prod	sum
5		X		prod	sum
6			X	prod	sum

Table 4: Errors defined for image ranking in case of descriptions on separate feature sets

Table 4 shows the definitions of the used errors. The third column 'weighted by train perf.' indicates that the ranking is also weighted by how well the extracted indices from the training image are accepted by the One-Class classifiers. When all indices of the training image are rejected by the SVDD's, the corresponding feature set gets a low weight. The difference between a sum or a product over feature sets is, that the sum performs an OR-operation and the product an AND-operation. In the cathedral query this means the difference between 'blue sky OR brick wall' and 'blue sky AND brick wall'.

In the fifth and sixth column of table 2 the results for the separate SVDD's without and with scaling is shown. The separate feature sets are scaled well and the rescaling does not improve performance very much. The results are somewhat worse for the Cathedral, Queen and Newsreader with respect to the original results. Only when the error contains a product over the feature sets (errors 4,5,6), performance is comparable.

Figure 6: Image of Hut, with one of the panels partly occluded by an antenna.

For the Hut image the OR-operation over the feature sets performs very well. It appeared that in some of the desired images one of the panels is partly obscured by an antenna, thus destroying the texture characterization of the original panel. An example is shown in figure 6. By using the OR-operation, the influence of the wrongly characterized panels can be reduced.

Finally the Flag query still suffers from the same problems as before. The two desired images which are rotated 90 degrees are not ranked very high. Best performance is reached using an AND-operation on the feature sets, while only considering the distance to the centers to the SVDD's.

In the last column of table 2 the best results on this database is shown, obtained by Messer[Mes99] on the same queries. Here a neural network is trained on a normal (two class) classification problem. Samples from the target classes are drawn from the user defined regions, but furthermore

outliers are drawn randomly from images from the rest of the database. An automatic training and pruning procedure is applied to this dataset to obtain the smallest and best performing neural network for the query at hand. In general the results are better, especially for the difficult queries. Drawback is that is requires a good set of outlier objects. It is not clear if randomly drawing objects from the database will always give a representative dataset (especially when a large dataset is used). Using a One-Class classifier avoids this problem.

5 Conclusion

The Support Vector Data Description is made to separate one class of data from the rest of the feature space. The method depends on well defined distances between objects and to investigate the influence of ill-defined distances in the data, the SVDD is applied to the problem of image database retrieval. In this problem objects are characterized by color and texture feature with very different scales.

It appeared that in this application the color feature is an important feature and that it is well clustered. Scaling all features to unit variance, both color and texture features, improves performance for image queries in which color features are most important. For queries where texture features are more important, the different color and texture features are better separated and treated separately. Combining the descriptions from the different feature sets opens the possibility of applying AND and OR operations on the different features. Then a distinction between 'blue sky AND brick wall' and 'blue sky OR brick wall' can be made. Of course this will require user input.

A normal operation procedure might therefore be, that an SVDD is trained on all 33 features, which are scaled to unit variance. In most cases this gives very acceptable results, especially when queries are focussed on color. When the user is not satisfied, data descriptions in separate feature sets should be trained. The user then has to indicate how they should be combined (by AND or OR combinations).

Of course this application can be improved in several ways. For instance, the user should be able to reject some images. These images than can be used as negative examples for the Support Vector Data Description in the next query. This is a subject for further research.

6 Acknowledgements

We would like to thank Kieron Messer for making available the image database with all preprocessed images. This work was partly supported by the Foundation for Applied Sciences (STW) and the Dutch Organisation for Scientific Research (NWO).

References

[AKJ98] S. Antani, R. Kasturi, and R. Jain. Pattern recognition methods in image and video databases: past, present and future. In *Advances in Pattern Recognition, Proceedings of SPR'98 and SSPR'98*, pages 31–53, Berlin, 1998. IAPR, Springer-Verlag.

[Mes99] K. Messer. *Automatic image database retrieval system using adaptive colour and texture descriptors.* PhD thesis, University of Surrey, Guildford, September 1999.

[TD98] D.M.J. Tax and R.P.W Duin. Outlier detection using classifier instability. In Amin, A., Dori, D., Pudil, P., and Freeman, H., editors, *Advances in Pattern Recognition, Lecture notes in Computer Science*, volume 1451, pages 593–601, Berlin, August 1998. Proc. Joint IAPR Int. Workshops SSPR'98 and SPR'98 Sydney, Australia, Springer.

[TD99] D.M.J. Tax and R.P.W Duin. Support vector domain description. *Pattern Recognition Letters*, 20(11-13):1191–1199, December 1999.

[TdRD97] D.M.J. Tax, D. de Ridder, and R.P.W. Duin. Support vector classifiers: a first look. In *Proceedings ASCI'97*. ASCI, 1997.

[Vap95] V. Vapnik. *The Nature of Statistical Learning Theory.* Springer-Verlag New York, Inc., 1995.

[Vap98] V.N. Vapnik. *Statistical Learning Theory.* Wiley, 1998.