

Bag Classification Using Support Vector Machines

Uri Kartoun, Helman Stern, Yael Edan

{kartoun|helman|yael}@bgu.ac.il

Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Be'er-Sheva, Israel

Abstract:

This paper describes the design of multi-category support vector machines (SVMs) for classification of bags. To train and test the SVMs a collection of 120 images of different types of bags were used (backpacks, small shoulder bags, plastic flexible bags, and small briefcases). Tests were conducted to establish the best polynomial and Gaussian RBF (radial basis function) kernels. As it is well known that SVMs are sensitive to the number of features in pattern classification applications, the performance of the SVMs as a function of the number and type of features was also studied. Our goal here, in feature selection is to obtain a smaller set of features that accurately represent the original set. A K-fold cross validation procedure with three subsets was applied to assure reliability. In a kernel optimization experiment using nine popular shape features (area, bounding box ratio, major axis length, minor axis length, eccentricity, equivalent diameter, extent, roundness and convex perimeter), a classification rate of 95% was achieved using a polynomial kernel with degree six, and a classification rate of 90% was achieved using a RBF kernel with 27 sigma. To improve these results a feature selection procedure was performed. Using the optimal feature set, comprised of bounding box ratio, major axis length, extent and roundness, resulted in a classification rate of 96.25% using a polynomial kernel with degree of nine. The collinearity between the features was confirmed using principle component analysis, where a reduction to four components accounted for 99.3% of the variation for each of the bag types.

Keywords: Support vector machines, multi-category classification, feature selection

1 Introduction

SVMs have emerged as very successful pattern recognition methods in recent years [16]. SVMs have yielded superior performance in various applications such as; text categorization [15], and face detection [12], content-based image retrieval [6], and learning image similarity [4].

The motivation here is detection of suspicious bags in a security situation. The usual method for bomb personnel is to blow up a suspicious bag, and any explosives contained therein. However, if the bag contains chemical, biological or radiological canisters, this may lead to disastrous results. Furthermore, the “blow-up” method also destroys important clues such as fingerprints, type of explosive, detonators and other signatures of importance for forensic analysis. Extraction of the bag contents using telerobotics avoids these problems [8]. In a telerobotic system, it is advantageous to automate bag classification which is coupled to robotic tactics such as shaking out of the bags contents. For cooperative human-robot interaction, in situations when autonomous capabilities fail, a human may be called in to distinguish the bag type. Here we focus on the autonomous operation of such a telerobotic system. In the autonomous mode, bags are classified by type using SVMs for the purpose of identifying initial manipulator grasp points. One side of the gripper must slide under the bag, and because of slippage, the grasp may fail. Thus, it is of importance to realize the type of bag, and the location of the opening before the grasp point assessment is made. Moreover, because we are dealing with a soft object here, with an unknown skin texture and unknown objects inside of the bag grasp may fail. Also, for each bag type a different rule set is evoked to determine the robot arm shake trajectories to discharge the contents of the bag for subsequent inspection.

In this paper we report on multi-category support vector classification of bags. The classification approach is presented in Section 2. Image processing operations are described in Section 3. Kernel optimization and optimal feature selection experimental results are described in Section 4. In section 5 an analysis of the results is given. The paper ends with conclusions and some directions for future work.

2 Classification Approach

SVMs belong to the class of maximum margin classifiers [5]. The goal of maximum margin classification in binary SVM classification [16] is to separate the two classes by a hyperplane such that the distance to the support vectors is maximized. SVMs perform pattern recognition between two classes by finding a decision surface that has maximum distance to the closest points in the training set which are termed support vectors. The procedure starts with a training set of points $x_i \in \mathfrak{X}^n$, $i = 1, 2, \dots, N$ where each point x_i belongs to one of two classes identified by the label $y_i \in \{-1, 1\}$. This hyperplane is called the optimal separating hyperplane (OSH). The OSH has the form:

$$f(x) = \sum_{i=1}^N \alpha_i y_i x_i \cdot x + b. \quad (2.1)$$

The coefficients α_i and b in (2.1) are the solutions of a quadratic programming problem [16]. A data point x is classified by the sign of the right side of (2.1). For multi-category classification, (2.2) is used.

$$d(x) = \frac{\sum_{i=1}^N \alpha_i y_i x_i \cdot x + b}{\| \sum_{i=1}^N \alpha_i y_i x_i \|}. \quad (2.2)$$

The sign of d is the classification result for x , and $|d|$ is the distance from x to the hyperplane. Intuitively, the farther away a point is from the decision surface, i.e., the larger $|d|$, the more reliable the classification result. The entire construction can be extended to the case of nonlinear separating surfaces. Each point x in the input space is mapped to a point $z = \Phi(x)$ of a higher dimensional space, called the feature space, where the data are separated by a hyperplane. The key property in this construction is that the mapping $\Phi(\cdot)$ is subject to the condition that the dot product of two points in the feature space $\Phi(x) \cdot \Phi(y)$ can be rewritten as a kernel function $K(x, y)$. The decision surface has the equation:

$$f(x) = \sum_{i=1}^N \alpha_i y_i K(x, x_i) + b, \quad (2.3)$$

where, the coefficients α_i and b are solutions of a quadratic programming problem. Note, that $f(x)$ does not depend on the dimensionality of the feature space. Kernel functions commonly used for pattern recognition problems are polynomial of degree d (2.4) and gaussian radial basis (2.5).

$$K(x, y) = (1 + x \cdot y)^d, \quad (2.4)$$

$$K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}. \quad (2.5)$$

The dual category SVM classification was extended to multi-category classification by [1]. There are two basic strategies for solving q -class problems with SVMs. In the one-vs-all approach, q SVMs are trained. Each of the SVMs separates a single class from all remaining classes [14, 2]. In the pairwise approach (used in this work), $q(q-1)/2$ machines are trained. Each SVM separates a pair of classes. The pairwise classifiers are arranged in trees, where each tree node represents a SVM. Regarding the training effort, the one-vs-all approach is preferable since only q SVMs have to be trained compared to $q(q-1)/2$ SVMs in the pairwise approach. The run-time complexity of the two strategies is similar: the one-vs-all and the pairwise approaches require the evaluation of q and $q-1$ SVMs,

respectively. Results on person recognition indicate similar classification performance for the two strategies [11]. The input to the SVMs is a set of features obtained from the bag image. The image processing feature extraction methods are described in the next section.

3 Image Processing

Image processing starts with a 24-bit color image of the robotic scene that contains a bag located on a platform. Four different bag types are considered: backpacks, small shoulder bags, plastic flexible bags, and small briefcases (Fig. 3.1). Image processing operations used are [7]: conversion to gray-scale, thresholding using the Otsu method [13], removal of noise from the image due to optical lens distortion, adaptation to ambient and external lighting conditions, and segmentation of the bag from the background. Using the MATLAB's Image Processing Toolbox [10], nine popular shape features were extracted from each segmented bag image: Area, Bounding Box Ratio, Major Axis Length, Minor Axis Length, Eccentricity, Equivalent Diameter, Extent, Roundness, Convex Perimeter.

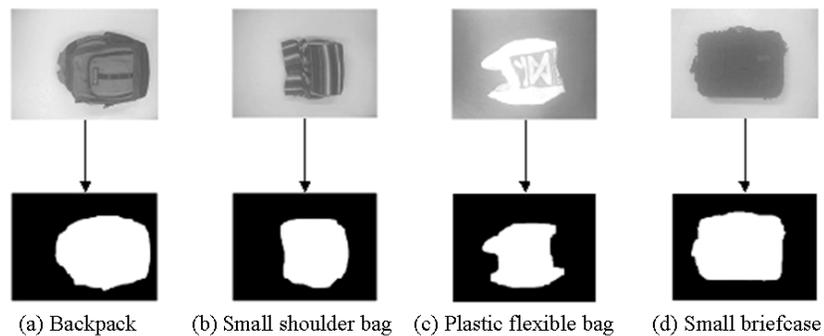


Fig. 3.1. Image processing operations

4 Experimental Results

4.1 Methodology

The heart of the support vector machine design is the kernel selection. Kernels project the data into a high dimensional feature space and thereby increase the

computational power of the linear learning machines [16, 3]. The kernels chosen for the bag classification problem are the most common ones used in support sector machines, i.e., the polynomial and the Gaussian RBF kernels [3]. The kernels were tested using the K-fold cross validation procedure with three subsets on a training set that contained 80 bag images (20 of each of the four classes). For testing, 40 bag images (10 of each class) were used. The SVMs procedure was implemented in the “OSU SVM MATLAB Toolbox” [9].

Two experiments were performed. In the first one, described in Section 4.2, a kernel optimization procedure was conducted, using the nine bag features described in Section 3, for finding the optimal polynomial degrees and RBF sigmas. In the second experiment, described in Section 4.3, an optimal feature selection was conducted for finding the number and combination of features that results in the highest classification rate using kernel optimization.

4.2 Kernel Optimization Experiment for Bag Classification

Classification process was performed for the range of 1-100 degrees / sigmas applying both kernels for finding the optimal polynomial degrees and RBF sigmas using all the nine features mentioned in Section 3.

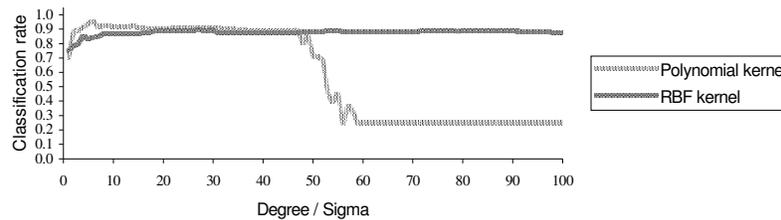


Fig. 4.2.1. Classification rate vs. degree and sigma for four bags classification using nine image features

As can be seen in Fig. 4.2.1 for the nine bag features case, the highest average classification rate between the three subsets achieved was 95% using a polynomial kernel with six degree and 90% using a RBF kernel with 27 sigma. The confusion matrices that summarize the results are shown in Table 4.2.1, where the upper and lower diagonal values in each cell correspond to percent of correct classifications for the polynomial and RBF kernels, respectively.

Table 4.2.1. Confusion matrices for polynomial/RBF kernels for a four bag classification using nine image features

predicted class	true class classification rates [%]			
	backpack	small shoulder bag	plastic flexible bag	small briefcase
backpack	96.7 / 93.3	0 / 0	0 / 0	3.3 / 6.7
small shoulder bag	3.3 / 0	93.4 / 93.3	0 / 0	3.3 / 6.7
plastic flexible bag	0 / 3.3	0 / 3.3	96.7 / 93.4	3.3 / 0
small briefcase	3.3 / 20	3.3 / 0	0 / 0	93.4 / 80

4.3 Optimal Feature Selection

A full enumeration feature selection procedure was performed to choose a set of optimal features to improve the results achieved in Section 4.2. Since there are up to nine features to be used in the classification procedure, there are $2^9 - 1 = 511$ combinations for selection. The classification process was performed for the range of 1-100 degrees / sigmas applying both kernels to find the optimal set of bag image features corresponding to the optimal polynomial degrees and RBF sigmas giving the highest classification results. As can be seen from Fig 4.3.1, the largest average classification rate can be obtained by using only four features for both the polynomial and RBF kernels.

Details of the feature selections and the average classification rates appear in Table 4.3.1. The polynomial kernel's average classification rate of 96.25% was superior to that of the RBF (91.6%). It is noted that for the polynomial kernel, the minimum set of features (among the ties) consists of four features: bounding box ratio, major axis length, extent and roundness. Also, from Table 4.3.1 these features correspond to the optimal polynomial kernel of degree nine.

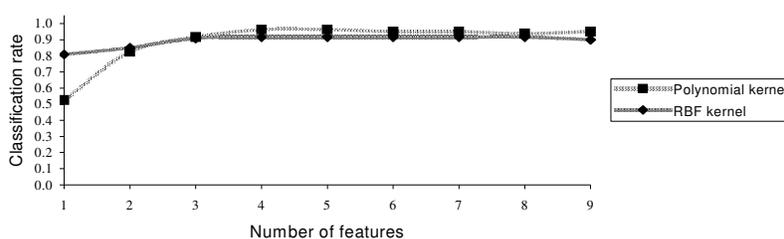


Fig. 4.3.1. Classification rate as a function of number of features using the optimal degrees / sigmas

Table 4.3.1. Feature selection results for the highest classification rates achieved

kernel number of features	polynomial			RBF		
	feature(s)	degree	classifi- cation rate [%]	feature(s)	sigma	classifi- cation rate [%]
1	G	6	52.5	F	53	80.8
2	G, I	6	82.5	A, G	9	85
3	D, F, G	8	91.6	D, G, H	33	90.8
4	B, C, G, H	9	96.25	B, D, E, G	48	91.6
5	B, E, F, G, H	6	96.25	B, D, E, G, H	39	91.6
6	B, C, D, E, G, H	6	95	B, C, D, E, G, H	60	91.6
7	B, C, D, E, G, H, I	6	95	B, C, D, E, G, H, I	56	91.6
8	A, B, C, D, E, F, H, I	6	93.75	A, C, D, E, F, G, H, I	63	91.6
9	A, B, C, D, E, F, G, H, I	6	95	A, B, C, D, E, F, G, H, I	27	90

A Area, B Bounding Box Ratio, C Major Axis Length, D Minor Axis Length, E Eccentricity, F Equivalent Diameter, G Extent, H Roundness, I Convex Perimeter.

Another view of the results are depicted in Fig. 4.3.2, using the optimal set of features (bounding box ratio, major axis length, extent and roundness). The curve shows that the average classification rate peaks for an optimal polynomial kernel degree nine.

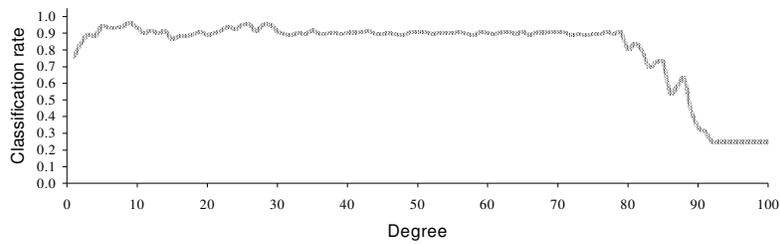


Fig. 4.3.2. Classification rate vs. degree for four-bag classification using optimal features (bounding box ratio, major axis length, extent and roundness)

5 Analysis of the Results

The polynomial kernel's classification rate of 96.25% was superior to that of the RBF (91.6%) using only four out of the nine original features. The small number of features deemed to be optimal was hypothesized to be due to correlation between the features. This is verified by examining the correlation matrix (Table 5.1), which exhibited correlation coefficients as high as 0.99.

Table 5.1. Correlation matrix

features	A	B	C	D	E	F	G	H	I
A	1	0.08	0.96	0.92	0.28	0.99	-0.27	-0.26	0.85
B	0.08	1	0.28	-0.18	0.69	0.1	0.01	-0.74	0.48
C	0.96	0.28	1	0.81	0.49	0.96	-0.3	-0.49	0.96
D	0.92	-0.18	0.81	1	-0.08	0.93	-0.31	0.11	0.61
E	0.28	0.69	0.49	-0.08	1	0.26	0	-0.99	0.69
F	0.99	0.1	0.96	0.93	0.26	1	-0.24	-0.25	0.86
G	-0.27	0.01	-0.3	-0.31	0	-0.24	1	0.02	-0.2
H	-0.26	-0.74	-0.49	0.11	-0.99	-0.25	0.02	1	-0.7
I	0.85	0.48	0.96	0.61	0.69	0.86	-0.2	-0.7	1

A Area, B Bounding Box Ratio, C Major Axis Length, D Minor Axis Length, E Eccentricity, F Equivalent Diameter, G Extent, H Roundness, I Convex Perimeter.

To further confirm the collinearity between the features we performed a principle component analysis (PCA), where a reduction to four components accounted for 99.3% of the variation for each of the bag types. Eigenvalues of the covariance matrix and the cumulative percentage of the total variance in the observations explained by the eigenvectors are shown in Table 5.2.

Table 5.2. Eigenvalues and variance explained matrix

variance explained [%]	62.65	93.46	96.9	99.31	99.79	99.93	99.97	100	100
eigenvalue	0.0948	0.056	0.009	0.003	0.0007	0.0003	0.0001	0	0

It is noted, that the reduction from nine to four features was done by a complete enumeration feature selection method. It is a coincidence that a PCA allowed a reduction to four "components", but the PCA was not used to represent the feature reduction through a linear combination of the original features, as in the usual case.

6 Conclusions

Multi-category bag classification for four-bag classes was performed using SVMs. The SVMs procedure was tested using polynomial and RBF kernels. A K-fold cross validation procedure with three subsets was used. In a kernel optimization experiment using nine popular shape features, classification rates of 95% and 90% were achieved using a polynomial kernel of degree six and a RBF kernel with 27 sigma, respectively. The confusion matrices indicate that decreased classification rates may be due to the inability to discriminate between the backpacks and the small briefcases bag images. To improve these results, a full enumeration feature selection procedure for choosing a set of optimal features for describing a bag image was performed. The set of optimal features found was: bounding box ratio, major axis length, extent and roundness. Using these features a classification rate of 96.25% was obtained for a polynomial kernel of degree nine. It was also found that using more than four features resulted in no improvement. The resulting optimal reduction of features from nine to four was hypothesized to be due to correlation between the features. The collinearity between the features was confirmed using principle component analysis, where a reduction to four components accounted for 99.3% of the variation for each of the bag types. Future work includes a comparison of the SVM classifier with that of using a PCA classifier.

Acknowledgments

This project was partially supported by the Ministry of Defense MAFAT Grant No. 1102, and the Paul Ivanier Center for Robotics Research and Production Management, Ben-Gurion University of the Negev. The assistance of Prof. Ehud Menipaz is gratefully acknowledged.

References

- [1] Bennett, K.P. and Bredensteiner, E.J. (1999), "Multicategory classification by support vector machines", *Computational Optimizations and Applications*, vol. 12, pp. 53-79.
- [2] Cortes, C. and Vapnik, V. (1995), "Support vector networks", *Machine Learning*, vol. 20, pp. 1-25.
- [3] Cristianini, N. and Shawe-Taylor, J. (2003), *Support Vector Machines and other Kernel-based Learning Methods*, Cambridge University Press, Cambridge, U.K.
- [4] Guo, G.D., Li, S.Z., and Chan, K.L. (2000), "Learning similarity for texture image retrieval", *Proceedings of the European Conference on Computer Vision*, pp. 178-190.

- [5] Heisele, B., Ho, P., Wu, J., and Poggio, T. (2003), "Face recognition: component-based versus global approaches", *Elsevier Science Inc.*, New York, U.S.A., vol. 91, pp. 6-21.
- [6] Hong, P., Tian, Q., and Huang, T.S. (2000), "Incorporate support vector machines to content-based image retrieval with relevance feedback", *Proceedings of the International Conference on Image Processing*, vol. 3, pp. 750-753.
- [7] Kartoun, U. (2003), "A human-robot collaborative learning system using a virtual reality telerobotic interface", *Ph.D. Thesis Proposal, Department of Industrial Engineering and Management at the Ben-Gurion University of the Negev, Israel*.
- [8] Kartoun, U., Stern, H., and Edan, Y. (2004), "Virtual reality telerobotic system", *e-ENGDET 2004 4th International Conference on e-Engineering and Digital Enterprise Technology*, Leeds Metropolitan University Yorkshire, U.K.
- [9] Ma, J. and Ahal, S. (2003), OSU SVM Classifier Matlab Toolbox (ver. 3.00).
- [10] The MathWorks Inc., (1998), Matlab Software, Image Processing User's Guide.
- [11] Nakajima, C., Pontil, M., Heisele, B., and Poggio, T. (2000), "Person recognition in image sequences: the MIT espresso machine system", *IEEE Transactions on Neural Networks*.
- [12] Osuna, E., Freund, R., and Girosi, F. (1997), "Training support vector machines: an application to face detection", *Proceedings of Computer Vision and Pattern Recognition*, pp.130-136.
- [13] Otsu, N. (1979), "A threshold selection method from gray level histograms", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, pp. 62-66.
- [14] Schölkopf, B., Burges, C., and Vapnik, V. (1995), "Extracting support data for a given task", In U. Fayyad and R. Uthurusamy, editors. *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, AAAI Press, pp. 252-257.
- [15] Shanahan, J.G. and Roma, N. (2003), "Boosting support vector machines for text classification through parameter-free threshold relaxation", *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, pp. 247-254.
- [16] Vapnik, V. (1998), *Statistical Learning Theory*, JohnWiley and Sons, New York.