

Genetic Outlier Detection for a Robust Support Vector Machine

Heesung Lee¹ and Euntai Kim²

¹Department of Railroad Electrical and Electronics Engineering, Korea National University of Transportation, Gyeonggi-do, Korea.

²School of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea.



Abstract

Support vector machine (SVM) has a strong theoretical foundation and also achieved excellent empirical success. It has been widely used in a variety of pattern recognition applications. Unfortunately, SVM also has the drawback that it is sensitive to outliers and its performance is degraded by their presence. In this paper, a new outlier detection method based on genetic algorithm (GA) is proposed for a robust SVM. The proposed method parallels the GA-based feature selection method and removes the outliers that would be considered as support vectors by the previous soft margin SVM. The proposed algorithm is applied to various data sets in the UCI repository to demonstrate its performance.

Keywords: SVM, Robust SVM, Genetic algorithm, Support vectors, Outlier

1. Introduction

Support vector machine (SVM) was proposed by Vapnik et al. [1, 2]; it implements structural risk minimization [3]. Beginning with its early success with optical character recognition [1], SVM has been widely applied to a range of areas [4-6]. SVM possesses a strong theoretical foundation and enjoys excellent empirical success in pattern recognition problems and industrial applications [7]. However, SVM also has the drawback of sensitivity to outliers and its performance can be degraded by their presence. Even though slack variables are introduced to suppress outliers [8, 9], outliers continue to influence the determination of the decision hyperplane because they have a relatively high margin loss compared to those of the other data points [10]. Further, when quadratic margin loss is employed, the influence of outliers increases [11]. Previous research has considered this problem [8-10, 12-14].

In [12], an adaptive margin was proposed to reduce the margin losses (hence the influence) of data far from their class centroids. The margin loss was scaled based on the distance between each data point and the center of the class. In [13, 14], the robust loss function was employed to limit the maximal margin loss of the outlier. Further, a robust SVM based on a smooth ramp loss was proposed in [8]. It suppresses the influence of outliers by employing the Huber loss function. Most works have aimed at reducing the effect of outliers by changing the margin loss function; only a small number have aimed at identifying the outliers and removing them in the training set. For example, Xu et al. proposed an outlier detection method using convex optimization in [10]. However, their method is complex and relaxation is employed to approximate the optimization.

Received: Apr. 8, 2015
Revised : Jun. 23, 2015
Accepted: Jun. 25, 2015

Correspondence to: Euntai Kim
(etkim@yonsei.ac.kr)
©The Korean Institute of Intelligent Systems

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, a new robust SVM based on a genetic algorithm (GA) [15] is proposed. The proposed method locates the outliers among the samples and removes them from the training set. The basic idea of this SVM parallels that of genetic feature selection, wherein GAs locate the irrelevant or redundant features and remove them by mimicking natural evolution. In the proposed method, GA detects and removes outliers that would be considered as support vectors by the previous soft margin SVM.

The remainder of this paper is organized as follows. In Section 2, we offer preliminary information on GAs. In Section 3, we describe the proposed method. Section 4 details the experimental results that demonstrate the performance and our conclusions are presented in Section 5.

2. Genetic Algorithms

Genetic Algorithms (GAs) are engineering models obtained from the natural mechanisms of genetics and evolution and are applicable to a wide range of problems. GAs typically maintain and manipulate a population of individuals that represents a set of candidate solutions for a given problem. The viability of each candidate solution is evaluated based on its fitness and the population evolves better solutions via selection, crossover, and mutation. In the selection process, some individuals are copied to produce a tentative offspring population. The number of copies of an individual in the next generation is proportional to the individual’s relative fitness value. Promising individuals are therefore more likely to be present in the next generation. The selected individuals are modified to search for a global optimal solution using crossover and mutation. GAs provide a simple yet robust optimization methodology [16].

3. Genetic Outlier Selection For Support Vector Machines

In this section, a new outlier detection method based on a genetic algorithm is proposed. First, dual quadratic optimization is formulated in a soft margin SVM and support vector candidates are selected from the training set based on the Lagrange multiplier. Then, the candidates are divided into either support vectors or outliers using GA. Figure 1 presents the overall procedure for the proposed method.

Suppose that M data points $\{x_1, x_2, \dots, x_M\}$ ($x_i \in R^n$) are given, each of which is labeled with a binary class $y_i \in \{-1, 1\}$.

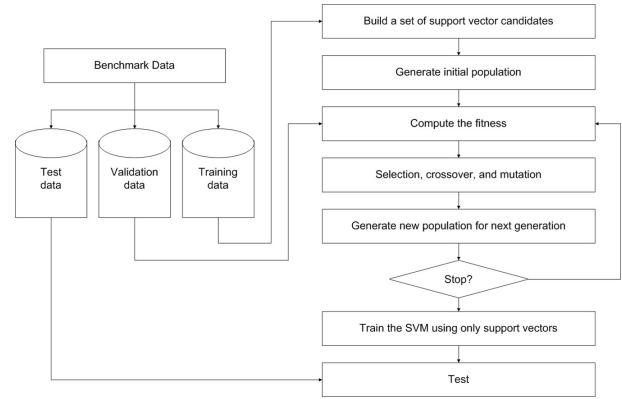


Figure 1. Procedure of the proposed method.

The goal of the SVM is to design a decision hyperplane

$$y_i = W^T x_i + w_0 \text{ for } i = 1, \dots, M \quad (1)$$

that maximally separates two classes

$$S_+ = \{x_j | y_j = 1\} \quad (2)$$

and

$$S_- = \{x_j | y_j = -1\}, \quad (3)$$

where W and w_0 are the weight and bias of the decision function, respectively. The SVM is trained by solving

$$\min_W \frac{1}{2} W^T W + C 1^T \Xi \quad (4)$$

subject to

$$Y (XW + Ew_0) - E + \Xi \geq 0 \quad (5)$$

$$\Xi \geq 0, \quad (6)$$

where $X = [x_1, x_2, \dots, x_M]^T$, $Y = \text{diag}(y_1, y_2, \dots, y_M)$, $1 = [1, 1, \dots, 1]^T$, and $0 = [0, 0, \dots, 0]^T$. $\Xi = [\xi_1, \dots, \xi_M]^T$ is a slack variable and implies a margin loss at each data point. C is a constant and denotes the penalty for a misclassification. The above formulation can be recast into

$$\max_{\Lambda} \Lambda 1 - \frac{1}{2} \Lambda Y X X^T Y \Lambda \quad (7)$$

subject to

$$1^T Y \Lambda = 0 \quad (8)$$

$$0 \leq \Lambda \leq C 1, \quad (9)$$

where $\Lambda = [\lambda_1, \lambda_2, \dots, \lambda_M]^T$ is a Lagrange multiplier vector and the nonnegative number λ_i is a Lagrange multiplier associ-

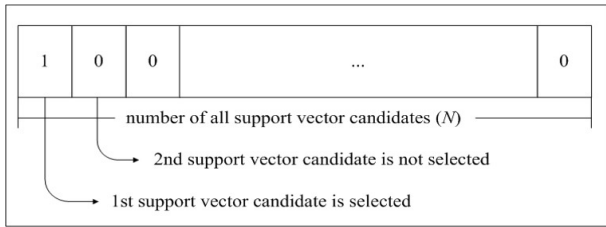


Figure 2. Chromosome used in the support vector selection.

ated with x_i . In a standard SVM, the data points with positive λ_i are support vectors and contribute to the decision hyperplane according to

$$y(x) = \sum_{i=1}^M \lambda_i y_i x_i^T x + w_0. \quad (10)$$

The interesting point is that if outliers are included in the training set, the outliers are likely to have positive margin loss and contribute to the hyperplanes. Further, the outliers tend to have relatively large margin loss and significantly influence the determination of the hyperplane, thereby making the SVM sensitive to the presence of outliers. In this paper, a robust SVM design scheme is proposed based on GA. First, a set of support vector candidates

$$S = \{x_i | \lambda_i > 0\} \quad (11)$$

is prepared by collecting the data points with positive Lagrange multipliers. As stated, not only the support vectors but also some outliers may be included in S . The goal of support vector selection is to determine a subset $S_v \subseteq S$ that includes only support vectors such that the classification accuracy of the SVM is maximized while the number of data points in the subset $card(S_v)$ is minimized, where $card(\cdot)$ denotes the cardinality. This is a bi-criteria combinatorial optimization problem and is usually intractable because it has an NP-hard search space. The implementation of the support vector selection parallels the feature selection method. The use of GA is a promising solution to this bi-criteria optimization problem because the feature selection methods based on GA outperform the non-GA feature selection methods [16-18]. To retain the support vectors and discard the outliers in subset S_v , the GA chromosome is represented by a binary string consisting of ones and zeros, as illustrated in Figure 2. In this figure, “1” and “0” indicate whether the associated data points should be retained or discarded in the set of support vectors, respectively. Genetic operators are applied to generate new chromosomes in the new generation. There are two types of genetic operators: crossover

Table 1. Experiment parameters

Parameter	Value
Crossover rate	0.6
Mutation rate	0.05
Population size	100
Fitness coefficient (α)	0.1
Control parameter for SVM (C)	100

Table 2. Datasets used in the experiments

	Number of instances	Number of attributes	Number of classes
Wine	178	13	3
Haberman	306	3	2
Transfusion	748	5	2
German	1000	20	2
Pima	768	8	2

and mutation. The purpose of the crossover is to exchange information among different potential solutions. The mutation introduces genetic material that may have been missing from the initial population or lost during crossover operations [19]. In this paper, one-point crossover and bit-flip mutation [20] are employed as genetic operators. When a validation set V is denoted as $V = \{v_1, v_2, \dots, v_m\}$, the fitness function of a chromosome is computed using

$$fitness = \frac{1}{m} \sum_{j=1}^m P_v(v_j) - \alpha(card(S_v)), \quad (12)$$

where

$$P_v(v_j) = \begin{cases} 1 & \text{If } v_j \text{ is correctly classified} \\ 0 & \text{otherwise} \end{cases}. \quad (13)$$

In this equation, m is the number of validation data points and α is a design coefficient. The fitness function actually implies the bi-criteria that the classification accuracy of the SVM should be maximized while the number of data points in the subset $card(S_v)$ should be minimized. The first term is aimed at improving the classification performance and the second term is aimed at the compactness of the SVM. The coefficient α plays an essential role in striking a balance between the classification performance and the classification cost. The parameters of the GA and SVM are given in Table 1.

In Table 1, α is set to 0.1 to emphasize the classification accuracy over the classification cost.

4. Experimental Results

In this section, the validity of the proposed scheme is demonstrated by applying it to five databases of the UCI repository [21]. The UCI repository has been widely used within the pattern recognition community as a benchmark problem for machine learning algorithms. The five databases are the Wine, Haberman, Transfusion, Garman, and Pima sets. All the sets except the Wine set are binary; first and second classes are used in the Wine set. The databases used in the experiments are summarized in Table 2.

In this experiment, the databases are randomly divided into four equal-sized subsets. Two subsets are used for training and the remaining two subsets are used for validation and testing. The training and validation sets are used to design a robust SVM and the test sets are used to evaluate the performance of the algorithms. To demonstrate the robustness of the proposed method against outliers, approximately 5% and 10% of the training samples were randomly selected from each class and their labels were reversed. Five independent runs were performed for statistical verification; the linear kernel was used for SVM. The performances of the proposed method and the general soft margin SVM were compared in terms of average testing accuracy and the number of support vectors. The results are summarized in Tables 3 and 4. In the tables, the proposed robust SVM is denoted as GASVM. It is observed that the standard SVM exhibits a marginally better performance than that of the proposed method for only the Australian database in the non-outlier case. In the majority of the cases, the proposed method achieves superior classification accuracy using a smaller number of support vectors than that of the standard SVM. That is, the proposed method is less sensitive to outliers and requires a reduced number of support vectors compared to the standard SVM. Further, by comparing the cases with 5% outliers and 10% outliers as indicated in Figure 3 and Figure 4, it can be observed that in the standard SVM, the greater the number of outliers that are included in the training set, the greater the number of support vectors generated and hence, the more the performance is degraded. In the proposed method, however, less sensitivity is exhibited toward the outliers and the increase in support vectors is limited. The reason for the improved performance of the proposed method is that only useful and discriminatory support vectors are selected and the brunt of the outlier influence on the SVM training is removed. To highlight the robustness of the proposed method, the test accuracy of the GASVM was normalized with respect to that

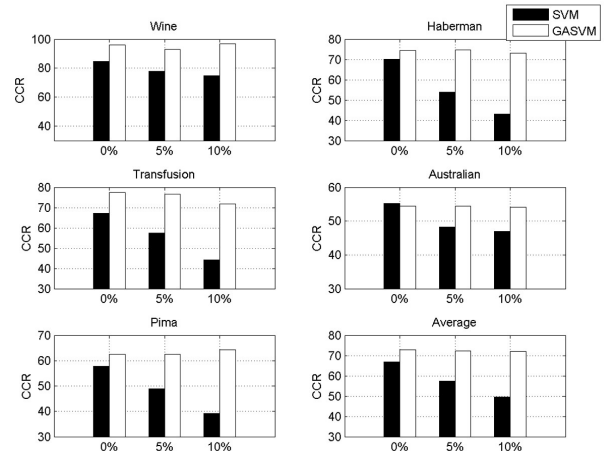


Figure 3. Correct classification ratio of the SVM and GASVM.

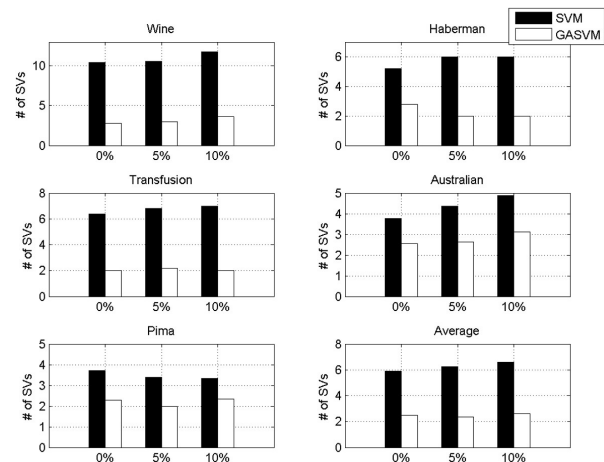


Figure 4. Number of support vectors of the SVM and GASVM.

of the standard SVM and the relative performances of the two SVMs are presented in Figure 5. In this figure, the length of the bar l denotes

$$l = C_{GASVM}/C_{SVM}, \tag{14}$$

where C_{GASVM} and C_{SVM} are the correct classification rates of GASVM and standard SVM, respectively. From this figure, it is clear that the greater the number of outliers included, the higher the relative excellence of the proposed method over the standard method.

5. Conclusions

In this paper, we presented a new method for detecting outliers to improve the robustness of SVM. The proposed method detected outliers within the support vectors assigned by soft

Table 3. Comparing the results of the proposed method (GASVM) with those of a previous method (SVM) in terms of testing accuracy

	No outliers			Including 5% outliers			Including 10% outliers		
	SVM	GASVM	Diff	SVM	GASVM	Diff	SVM	GASVM	Diff
Wine	84.50	96.00	11.50	78.00	93.00	15.00	75.00	96.67	21.67
Haberman	70.06	74.34	4.28	54.00	74.67	20.67	43.11	73.16	30.05
Transfusion	67.13	77.55	10.42	57.39	76.54	19.15	44.26	71.76	27.5
Australian	55.18	54.48	-0.7	48.25	54.42	6.17	46.92	54.08	7.16
Pima	57.78	62.44	4.66	49.01	62.60	13.59	39.12	64.24	25.12
Average	66.93	72.96	6.03	57.33	72.25	14.92	49.68	71.98	22.30

Table 4. Comparing the results of the proposed method (GASVM) with those of a previous method (SVM) in terms of the number of support vectors

	No outliers			Including 5% outliers			Including 10% outliers		
	SVM	GASVM	Diff	SVM	GASVM	Diff	SVM	GASVM	Diff
Wine	10.40	2.80	-7.60	10.60	3.00	-7.60	11.80	3.60	-8.20
Haberman	5.20	2.80	-2.40	6.00	2.00	-4.00	6.00	2.00	-4.00
Transfusion	6.40	2.00	-4.40	6.80	2.20	-4.60	7.00	2.00	-5.00
Australian	3.78	2.55	-1.23	4.37	2.63	-1.74	4.88	3.13	-1.75
Pima	3.71	2.28	-1.43	3.40	2.00	-1.40	3.33	2.33	-1.00
Average	5.90	2.47	-3.41	6.23	2.37	-3.87	6.60	2.61	-3.99

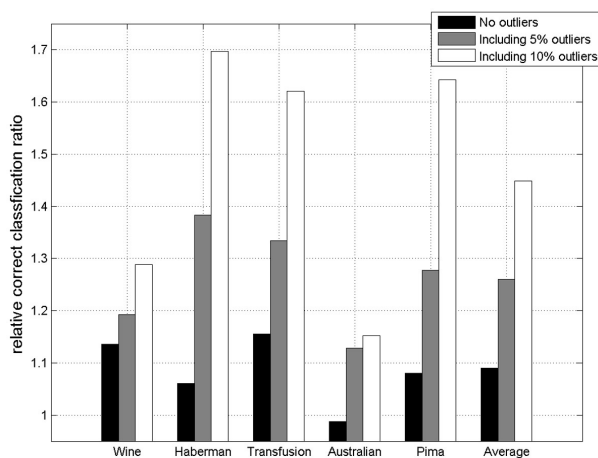


Figure 5. Relative performance of the proposed method compared to a general SVM.

margin SVM using GA, and demonstrated recognition performance and a total number of support vectors superior to those of previous methods. Using the proposed method, the robustness of SVM was improved and the SVM was simplified by outlier deletion. The validity of the suggested method was demonstrated through experiments with five databases from the UCI repository.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2013R1A2A2A01015624)

References

- [1] C. Cortes, and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, Sep. 1995.
- [2] V. N. Vapnik, *Statistical Learning Theory*. Wiley, 1998.
- [3] S. Jun, "An outlier data analysis using support vector regression," *Journal of The Korean Institute of Intelligent Systems*, vol. 18, no. 6, pp. 876-880, 2008.
- [4] V. Hoang, M. Le, and K. Jo, "Hybrid cascade boosting machine using variant scale blocks based HOG features for pedestrian detection," *Neurocomputing*, vol. 135, pp. 357-366, 2014.
- [5] S. Seo, H. Yang, K. Sim, "Behavior learning and evolution of swarm robot system using support vector machine,"

- Journal of The Korean Institute of Intelligent Systems*, vol. 18, no. 5, pp. 712-717, 2008.
- [6] H. Shin, H. Jung, K. Cho, and J. Lee “A prediction method of learning outcomes based on regression model for effective peer review learning,” *Journal of The Korean Institute of Intelligent Systems*, vol. 22, no. 5, pp. 624-630, 2012.
- [7] S. Kumar, *Neural Networks: A Classroom Approach*. McGraw-Hill, 2005.
- [8] L. Wang, H. Jia, and J. Li, “Training robust support vector machine with smooth ramp loss in primal space,” *Neurocomputing*, vol. 71, pp. 3020-3025, 2008.
- [9] H. Lee, S. Hong, B. Lee and E. Kim “Design of robust support vector machine using genetic algorithm,” *Journal of The Korean Institute of Intelligent Systems*, vol. 20, no. 3, pp. 375-379, 2010.
- [10] L. Xu, K. Crammer, and D. Schuurmans, “Robust support vector machine training via convex outlier ablation,” in *Proc. the 21st National Conference on Artificial Intelligence*, pp. 536-542, 2006.
- [11] J. A. K. Suykens, and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural Processing Letters*, vol. 9, no. 3, pp. 293-300, 1999.
- [12] Q. Song, W. Hu, and W. Xie, “Robust support vector machine with bullet hole image classification,” *IEEE Trans. Systems, Man, and Cybernetics-Part C: Applications and Reviews*, vol. 32, no. 4, pp. 440-448, 2002.
- [13] N. Krause and Y. Singer, “Leveraging the margin more carefully,” in *Proc. the 21st International Conference on Machine Learning*, vol. 69, 2004.
- [14] P. Bartlett and S. Mendelson, “Rademacher and Gaussian complexities: risk bounds and structural results,” *Journal of Machine Learning Research*, vol. 3, pp. 463-482, 2002.
- [15] L. Davis, *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, 1991.
- [16] H. Lee, E. Kim, and M. Park, “A genetic feature weighting scheme for pattern recognition,” *Integrated Computer-Aided Engineering*, vol. 14, pp. 161-171, 2007.
- [17] L. Kuncheva and L. Jain, “Nearest neighbor classifier: simultaneous editing and feature selection,” *Pattern Recognition Letters*, vol. 20, pp. 1149-1156, 1999.
- [18] I. Oh, J. Lee, and B. Moon, “Hybrid genetic algorithms for feature selection,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1424-1437, 2004.
- [19] H. Juo and H. Chang, “A new symbiotic evolution-based fuzzy-neural approach to fault diagnosis of marine propulsion systems,” *Artificial Intelligence*, vol. 17, pp. 919-930, 2004.
- [20] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, 1996.
- [21] P. M. Murphy and D. W. Aha, “UCI Repository for Machine Learning Databases,” *Technical report, Dept. of Information and Computer Science*, Univ. of California, Irvine, Calif., 1994.



Heesung Lee received the BS, MS, and PhD degrees in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2003, 2005, and 2010, respectively. From 2011 to 2014, he was a managing researcher with the S1 Corporation, Seoul, Korea. Since 2015, he has been with the railroad electrical and electronics engineering at Korea National University of Transportation Gyeonggi-do, Korea, where he is currently an assistant professor. His current research interests include computational intelligence, biometrics, and intelligent railroad system.



Euntai Kim received the BS (with top honors), MS, and PhD degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1992, 1994, and 1999, respectively. From 1999 to 2002, he was a full-time lecturer with the Department of Control and Instrumentation Engineering at Hankyong National University, Gyeonggi-do, Korea. Since 2002, he has been with the School of Electrical and Electronic Engineering at Yonsei University, where he is currently professor. He was a visiting scholar with the University of Alberta, Edmonton, Canada, in 2003, and is now a visiting researcher with the Berkeley Initiative in Soft Computing (BISC), UC Berkeley, USA. His current research interests include computational intelligence and machine learning and their application to intelligent service robots, unmanned vehicles, home networks, biometrics, and evolvable hardware.