A Method of Feature Selection Using Contribution Ratio based on Boosting

Masamitsu Tsuchiya, Hironobu Fujiyoshi
Dept. of Computer Science, Chubu University
Aichi 487-8501 Japan
{tsuchiya, hf}@vision.cs.chubu.ac.jp

Abstract

AdaBoost and support vector machines (SVM) algorithms are commonly used in the field of object recognition. As classifiers, their classification performance is sensitive to affected by feature sets. To improve this performance, in addition to using the classifiers for accurate selection of feature sets, attention must be given to determining which feature subset to use in the classifier. Evaluating feature sets using a margin of the decision boundary of an SVM classifier proposed by Kugler is a solution for this problem. However, the margin in an SVM is sometimes large due to outliers. This paper presents a feature selection method that uses a contribution ratio based on boosting, which is effective for evaluating features. By comparing our method to the conventional one that uses a confident margin, we found that our method can select better feature sets using the contribution ratio obtained from boosting.

1 Introduction

Feature-based classifiers are commonly used for object recognition and type classification [1]. To make the classifiers more robust, we need to select features that are invariant to various changes resulting from location, scaling, viewpoint, and other environments. To improve classification performance, in addition to using classifiers for accurate selection of feature sets such as AdaBoost [2], and SVM [3], determining which feature subset to use in the classifier is required.

Also, to reduce computation costs and maintain classification performance, removing ineffective features is important. Therefore, a tiny and effective feature subset selection (FSS) is needed for robustness and speed, particularly in image/object recognition algorithms that use large feature sets [5], [6]. Previous work in this area has focused on the margin of support vector machines (SVM). The problem with directly using the SVM mar-gin is that it does not always provide a clear relationship between its value and the performance of the SVM, and the best obtained subset is not guaranteed to be the best possible one. Kugler et. al. proposed FSS using the confident margin (CM) in the subset criterion, which enables FSS to come close to the best recognition rate by monitoring the peak of the CM curve without directly calculating the recognition rate, which reduces computational time [4]. However, in these methods, leaving a feature is necessary so that the margin of the SVM can be measured for as many number of features as there are in a feature subset. In addition, margin errors are present in the SVM in some cases.

This paper describes our novel method for selecting feature subsets using a contribution ratio (CR). The CR is the relative importance of features based on the boosting classifier for object type classification [8].

Starting with a given set of features, we estimate the CR of each features from a feature set and remove the features that are not contributing and that have a low CR until achieving the desired number of feature subsets. We evaluated our method by performing two tasks with a classification problem in a real world situation, and we confirmed that feature subset selection using the CR is more reliable than the method based on the margin of an SVM.

2 Related works[4]

When constructing an SVM, the distance between a decision boundary and a learning sample is maximized. This distance is called a "margin". The function of an SVM is determined by:

$$f(x) = \langle \mathbf{w} \cdot \Phi(\mathbf{x}) \rangle + b \qquad (1)$$

$$= \sum_{i=1}^{n} w_i \Psi(x) + b, \qquad (2)$$

where $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ is the input, $\mathbf{Y} = \{-1, 1\}$ is the label, $\mathbf{w}$ is the weight vector, Φ is the projection to the

nonlinear space, and b is the bias.

The margin of an SVM is represented as $\frac{1}{||\mathbf{w}||}$ by weight vector $\mathbf{w}$. The weight vector $\mathbf{w}$ is given by the follow equation, which solves equation 1 using the Lagrange multiplier method:

$$\mathbf{w} = \sum_{i=1}^{n} y_i \alpha_i \Phi(\mathbf{x}), \qquad (3)$$

where α is the Lagrange multipliers. A large margin classifier is expected to have a high generalization. Therefore, the margin can be used to evaluate the performance of the classifier and input features of an SVM.

The margin in an SVM, however, is sometimes large due to outliers [4]. In such cases, errors are present when using the margin-based method. Kugler et al. used the confident margin, which is a combination of confidence and margin (normal margin : NM). Confidence c is given by:

$$c = \frac{1}{l} \sum_{i=1}^{n} y_i f(\mathbf{x}_i). \qquad (4)$$

The defined value c has a stronger classification performance. The decision function $f(\mathbf{x}_i)$ of an SVM that can be expressed by the equations 1 and 3 as:

$$f(\mathbf{x}) = \sum_{i=1}^{n} y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b \qquad (5)$$

$$K(\mathbf{x}_i, \mathbf{x}) = \langle \Psi(x_i) \cdot \Psi(x) \rangle. \qquad (6)$$

The confident margin CM, estimated by using confidence c and normal margin NM, is as follows:

$$CM = c \cdot NM. \qquad (7)$$

Using the confident margin, the variance between the fact performance and the evaluation can be reduced.

However, even the confident margin is not free from margin errors, as we later discuss. We have to deal with the problem of not havinge an outlier in a dataset.

3 Feature contribution ratio based on boosting

A feature subset evaluation using the margin of an SVM is problematic due to an error from outliers. Also, constructing the same number of SVM classifiers as the square numbers of features is needed.

We focused on a boosting classifier that is equal to an SVM to maximize the margin. The boosting classifier was constructed so that its features could be used more effectively. Tsuchiya et al. has proposed a feature evaluation method based on boosting [8]. They defined the feature contribution ratio with a weak hypothesis and the weight from the performance of the weak hypothesis.

3.1 AdaBoost

AdaBoost selects a subset of features to construct a robust classifier from a training dataset $\{(\boldsymbol{x}_i, y_i) : 1 < i < n\}$, where $\boldsymbol{x} = (x^1, \cdots x^P)$ is the 11-dimensional feature vector, and $y \in \{-1, +1\}$ is the label as follows:

$$y_i = \begin{cases} +1, if\ \boldsymbol{x}_i \in target \\ -1, otherwise. \end{cases} \qquad (8)$$

In each round, the learning algorithm selects from all the features. The AdaBoost algorithm picks the optimal threshold th for each feature p by:

$$h_t^{p,th}(\boldsymbol{x}) = \underset{(0\ \leq\ th\ \leq\ 1,\ 1\ \leq\ p\ \leq\ P)}{argmin} \{\sum_i I(y_i = \mathrm{sgn}\,(x_i^p - th))\} \qquad (9)$$

The output of Adaboost after the learning process is a binary classifier that consists of a linear combination of the selected features with weights α_t. Therefore, the final classifier H_T is given by:

$$H_T(\boldsymbol{x}) = \sum_{t=1}^{T} \alpha_t h_{(t)}(\boldsymbol{x}). \qquad (10)$$

3.2 Evaluating feature contribution ratio

In each round, Adaboost selects from the total set of various features, which are the features with the lowest weighted error on the training examples. The final classifier balances the 11 features to maximize classification performance. The weight α, the selected feature, and the threshold th chosen at each round are very important factors for bolstering classification performance.

Here, we introduce a metric that indicates how well the features "contribute" to the classification performance. A contribution ratio CR_p for each feature p is defined by:

$$CR_p = \sum_{t=1}^{T} \alpha_t' \cdot \delta_K[P(h_t) - p], \qquad (11)$$

where p is a kind of feature, and $P()$ is a function for outputting the feature chosen at round t in the AdaBoost training process. Let δ_K be the Kronecker delta. This contribution ratio CR_p becomes a metric for measuring the contribution of the feature vector p, and it enables us to determine which subset of features should be selected in a given classification task. This evaluating method need only an AdaBoost classifier for evaluating all the

features in the subset. In addition, CR is the invariant outlier than margin of an SVM, that caused by CR is estimated from performance and frequently in used of weak hypothesis.

We proposed a feature selection method using a contribution ratio based on boosting.

4 Feature subset selection using contribution ratio

A margin can be used to evaluate only a classifier on a feature subset, and it is not robust. Therefore, a margin-based method is problematic because constructing many classifiers is needed and margin is following no fixed the classification performance for outliers in some cases.

However, we can evaluate all the features in a subset using contribution ratio (CR) from an AdaBoost classifier. Thus, CR is an invariant outlier than margin of an SVM, and that necessitates the construction of as many AdaBoost classifiers as there are features. We developed a feature selection method using a contribution ratio based on boosting.

An illustration of our method is shown in Figure1. Starting with a given set of features, we estimate the CR of each features from a feature set and remove the features that are not contributing until achieving the number of desired feature subsets. Feature subset selection using a contribution ratio algorithm is given as shown in Figure2.
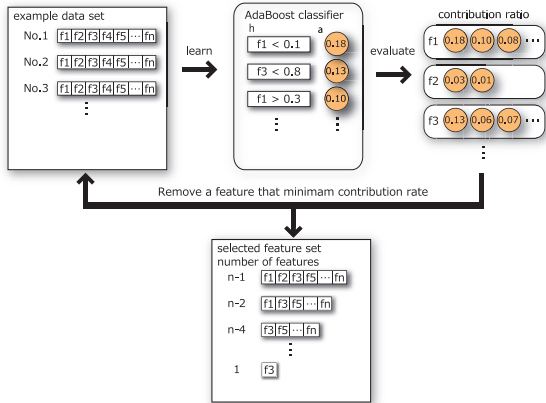

Figure 1. FSS using contribution ratio

5 Evaluation

We used two test sets that included binary classification problems called "Sonar" data from a UCI database

Algorithm The FSS using Contribution ratio
1. **Input:** n, Training dataset $(\mathbf{x}_i, y_i)$
2. **Initialize:**
Subset of surviving features $\boldsymbol{s} = [1, 2, ..., n]$
3. **Do for** Until $\boldsymbol{s}$ is empty
(a) Train AdaBoost classifier
with all the training exsamples
(b) Compute Contribution Rate
$\boldsymbol{CR}_1, \boldsymbol{CR}_2, ..., \boldsymbol{CR}_n$
(c) find the worst feature
$worst = argmin(\boldsymbol{CR}_i)$
(d) Remove the worst feature i that
minimam $\boldsymbol{CR}_i$

Figure 2. FSS Algorithm

[9] and "VH" data from a CU database [8]. The "Sonar" data included 208 samples with 60 features. The "VH" data included 800 samples with 7 features.

Our method is using only boosting classifiers. However, we believe this method has usable for common classifiers. To demonstrate we our hypothesis is real, we constructed an SVM classifier from selected feature subsets that are based on a confident margin and contribution ratio. Estimating the recognition rate using leave-one-out and cross-validation methods in order to estimate CM from separable datasets is difficult.

Our goal in these experiments is to show that, by comparing feature subset selection methods using CR and CM, we can improve the accuracy and effectiveness of selection from an outlier that includes errors related to CM. The SVM classifier and margin estimation are hindered by svm-lighit[7].

5.1 Sonar data case

We report the best recognition rate (Best RR) and feature dimension (DIM) using feature selection in table1.

Both methods are effective for reducing the number of features. Using CM is more effective than our method; however, our method can be used to select effective features that can maintain a recognition rate.

5.2 VH data case

We report the Best RR and DIM using feature selection in table2, and we show a variation of the recognition rate in figure 4.

Both methods are effective for reducing the number of features. However, the recognition rate using CM is 20% less effective than our method when selecting the last feature. We believe this problem is due to the SVM margin.
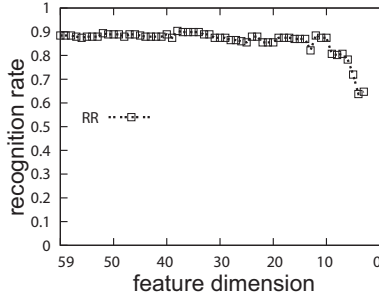
Figure 3. Variation of recognition rate with FSS on Sonar dataset

Table 1. Results of FSS on Sonar dataset

	Best RR[%]	DIM
Our method	91	36
Confident Margin	93	13
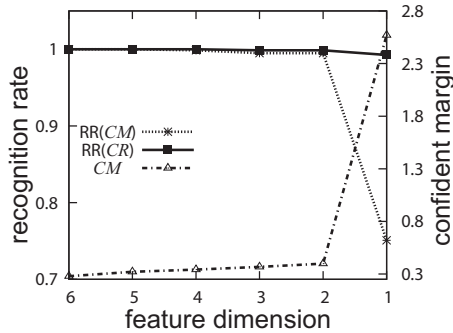
Figure 4. Variation of recognition rate with FSS on VH dataset

Table 2. Results of FSS on VH dataset

	Best RR[%]	DIM
Our method	100	3
Confident Margin	100	4

5.3 Margin -based evaluation problem

For the focus variation of confident margin (CM), shown in figure 4, the CM is very high when the last feature is selected. This is because for selection using the CM, ineffective features cannot be rejected.

The confidence (c) and normal margin (NM), when the last feature is selected, are shown in table 3. The confidence work to reduce margin in both cases, however, the NM is too large to effectively estimate the feature in the VH data. This means that the confidence has not been sufficiently checked. On the other hands, our method can effectively be used for selecting feature subsets. In addition, our method does not reject the most effective features, when the CR is used and does not affect the outlier at the margin of an SVM.

Table 3. Margin and confidence (DIM = 1)

	NM	c	CM
Sonar	0.57	0.12	0.06
VH	5.13	0.50	2.57

6 Conclusion

We created a novel feature subset selection method that uses a contribution ratio. The contribution ratio is obtained from the selected features and weights in the AdaBoost training.

We experimentally validated our method by demonstrating robust selection using an SVM classifier in a test set under an outlier. This enabled us to determine which feature should be selected, in a [general learning-based classifier.

References

[1] R. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade, "Algorithms for cooperative multisensor surveillance", Proc. of the IEEE, Vol. 89, No. 10, pp. 1456 - 1477 (Oct. 2001).

[2] Y. Freund and R.E. Schapire , "A decision-theoretic generalization of on-line learning and an application to boosting", Journal of Computer and System Sciences, pp. 119 -139, (1997)

[3] N. Cristianini, and J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, (2000.)

[4] M. Kuguler, K. Aoki, A. S. Nugroho, S. Kuroyanagi, and A. Iwata, "A Feature Subset Selection for Support Vector Machines using Confident Margin", IJCNN, Montreal., Proceedings, IEEE Computer Society, (2005.)

[5] D. Hoiem, A. A. Efros, and M. Hebert, "Putting Objects in Perspective", pp. 2137 - 2144, CVPR (2) (2006).

[6] P. Sabzmeydani and G. Mori, "Detecting Pedestrians by. Learning Shapelet Feature", CVPR, pp. 511- 518, (2007).

[7] T. Joachims," Making large-Scale SVM Learning Practical: Advances in Kernel Methods - Support Vector Learning ", B. Scholkopf, C. Burges, and A. Smola (ed.) MIT-Press (1999).

[8] M. Tsuchiya, and H. Fujiyoshi, "Evaluating Feature Importance for Object Classification in Visual Surveillance", Proc. of ICPR, pp. 978 - 981, 4 pages (2006).

[9] C. Blake, E. Keogh, and C. J. Merz, UCI Repository of machine learning databases [http://www.ics.uci.edu/mlearn/MLRepository.html]., Irvine, CA: University of California, Department of Information and Computer Science, (1998).