Feature Selection for Image Classification Based on a New Ranking Criterion

Xuan Zhou, Jiajun Wang

School of Electronic and Information Engineering, Soochow University, Suzhou, China
Email: jjwang@suda.edu.cn

Abstract

In this paper, a feature selection method combining the reliefF and SVM-RFE algorithm is proposed. This algorithm integrates the weight vector from the reliefF into SVM-RFE method. In this method, the reliefF filters out many noisy features in the first stage. Then the new ranking criterion based on SVM-RFE method is applied to obtain the final feature subset. The SVM classifier is used to evaluate the final image classification accuracy. Experimental results show that our proposed relief- SVM-RFE algorithm can achieve significant improvements for feature selection in image classification.

Keywords

Image Classification, Feature Selection, Ranking Criterion, ReliefF, SVM-RFE

1. Introduction

Feature selection plays a key role in many pattern recognition problems such as image classification [1] [2]. While a great many of features can be utilized to characterize an image, only a few number of them are efficient and effective in classification. More features does not always lead to a better classification performance, thus feature selection is usually performed to select a compact and relevant feature subset in order to reduce the dimensionality of feature space, which will eventually improves the classification accuracy and reduce time consumption [2].

Based on different evaluation criterion, feature selection methods can be classified to two categories: filter models and wrapper models. Filter models generally utilize the characteristics of feature data and are more computationally efficient. The reliefF algorithm is a typical example of this type which can effectively provide quality of each feature in problems with dependencies among the feature space [3]. It has been shown by Zhang and Ding that the reliefF algorithm performs well in gene selection [4].

As compared with the filter type selectors, the wrapper models can usually achieve higher classification ac-

curacy. Such superiority is achieved by involving the classifier in the selection process. The SVM recursive feature elimination (SVM-RFE) method is a typical wrapper type selector, where the support vector machine (SVM) [5] is used as a classifier. This method was firstly employed in gene selection [6], where the ranking criterion for different features is determined as the magnitude of the weights of the trained SVM classifier. Although SVM-RFE method always performs better in classification accuracy, it is usually computationally much more intensive [7].

In this paper, a new hybrid model is proposed by combing reliefF algorithm and SVM-RFE method. The new method (we call it reliefF-SVM-RFE) not only performs better than either of the two methods but also costs much less computational time than the wrapper models. The integration of reliefF and SVM-RFE leads to an effective feature selection scheme. In the first stage, reliefF is applied to find a subset of candidate features so that many irrelevant features are filtered out and the computational demands can be reduced. In the second stage, SVM-RFE method is applied to directly estimate the quality of each feature resulted from the reliefF algorithm according to our proposed evaluation criterion. Comprehensive experiments are performed to compare our new reliefF-SVM-RFE selection algorithm with the reliefF and SVM-RFE methods on our dataset picked from Caltech-256 database [8]. The experimental results demonstrate that the proposed feature selection algorithm outperforms the other two methods.

2. Feature Extraction

In order to obtain an effective feature subset by feature selection, the original feature set must be sufficient. Since low level visual features such as color, texture, and shape are fundamental to characterize images [9]-[11], 75 features of these three types are extracted to compose the pool of features for selection.

Totally 54 color descriptors are extracted from each image by combining different color models and quantization strategies. Texture descriptors characterize the structural pattern of an image, therefore 11 texture descriptors are studied in total. Similar to color and texture descriptor, shape feature is also an important low-level descriptor and has been widely used in image classification. Totally 10 shape descriptors are collected from each image.

Note that most of the features have more than one dimension and all 75 feature descriptors consist of 3268-dimension feature components. In the next analysis, each feature component will be treated as a single feature to be selected by our methods.

3. Feature Selection

3.1. reliefF Algorithm for Feature Selection

ReliefF is a simple yet efficient procedure to estimate the quality of feature in problems with strong dependencies between attributes [4]. In practice, reliefF is usually applied in data pre-processing for selecting a feature subset.

The key idea of the reliefF is to estimate the quality of feature according to how well their values distinguish between instances that are near to each other. reliefF algorithm randomly selects an instance R_i from its class, and then searches for K of its nearest neighbors from the same class, called nearest hits H, and also K nearest neighbors from each of the different classes, namely nearest misses M. Then it updates the quality measure W_i for feature i according to values for R_i, hits H and misses M. If instance R_i and those in H have different quality estimation on feature i, then the estimation W_i is decreased. Meanwhile, if instance R_i and those in M have different values on feature i, the estimation W_i will be increased. The process is repeated n times. Here, n is usually set by users and for most purposes it can be safely set to 10 [3]. According to the aforementioned discussion, the quality measure W_i can be updated as follows:

$$W_i = W_i - \frac{\sum_{k=1}^{K} D_H(k)}{n \cdot k} + \sum_{c=1}^{C-1} p_c \cdot \frac{\sum_{k=1}^{K} D_M(k)}{n \cdot k} \tag{1}$$

where $D_H(k)$ (or $D_M(k)$) is the sum of distance between the selected instance and its kth nearest neighbor in H (or M), p_c is the prior probability of class c. **Figure 1** gives a detailed description of the reliefF algorithm.

```
Input：Feature data matrix: D，repeat times: n，the number of the neighbors: K
Output：Vector W for the feature attributes ranking
Begin
    for j=1 to n do
        Randomly select an instance Rj;
        Find K nearest hits H and nearest misses M;
        for i=1 to all features do
            Updating estimation Wi by Equation(1);
        end
    end
End
```

Figure 1. Pseudo-code of reliefF algorithm.

3.2. SVM-RFE Algorithm for Feature Selection

The SVM-RFE for feature selection is a feature elimination method based on SVM classifier and capable of selecting a small feature subset [12]. It starts with all the feature components and recursively removes the feature components with the least importance for the classification in a backward elimination manner. In the original SVM-RFE method, the ranking criterion is computed from the weight vector of SVM as follows:

$$w = \sum_l \alpha_l \gamma_l \chi^{(l)} . \tag{2}$$

where α_l is the Lagrange multipliers, γ_l is the class label of the l th sample, and $\chi^{(l)}$ is the feature vector. With this weight vector, the importance of the i th feature component can be determined as $C_i = w_i^2$, and hence the feature component can be selected according to their ranking criterion above.

3.3. An Improved Algorithm by Combing reliefF and SVM-RFE

As we mentioned before, reliefF is a general and successful feature attributes estimator and is able to effectively provide quality estimation of features in problem with dependencies between feature attributes. However reliefF does not explicitly consider the role of classifier in feature extraction. On the other hand, the SVM-RFE algorithm totally takes the importance of classifier into account, it is computationally expensive [7]. Thus it could be helpful to integrate the weight vector W_i from reliefF into SVM-RFE method. Therefore, we devise a new ranking criterion as follows:

$$G_i = W_i + C_i . \tag{3}$$

where C_i is the ranking indicator from SVM-RFE method. It should be pointed out that before the sum operation of W_i and C_i those two weight vectors should be normalized to [0, 1], which can eliminate the error caused by two different dimensions. Thus the G_i is the final ranking indicator for the i th feature in our image classification.

In our experiments, reliefF is first used to select 300 features as the candidate subset from the all feature data, SVM-RFE is then utilized to choose the final subset according to our new ranking criterion.

4. Experimental Results and Discussion

Our experimental dataset, named dataset-Caltech, includes six categories of images which is randomly picked from Caltech-256 database [8]. Half of the images of each category are used for training and the others for testing. Representative samples of dataset-Caltech are shown in **Figure 2** and the detail description of our experimental dataset can be found in Caltech-256 database. In this section, comprehensive experiments are performed on dataset-Caltech to compare our new proposed selection algorithm with the reliefF and SVM-RFE algorithms.

In the experiment on dataset-Caltech, all the 75 feature descriptors as discussed above are first extracted for each image from dataset-Caltech, and then a 3268-dimension feature space can be obtained. In the second step, we consider each feature component in the feature space as a single feature, finally a compact and effective subset with about 150 feature components will be obtained. **Figure 3(a)** and **Figure 3(b)** show the classification accuracy achieved when different number of feature components are selected for classification.
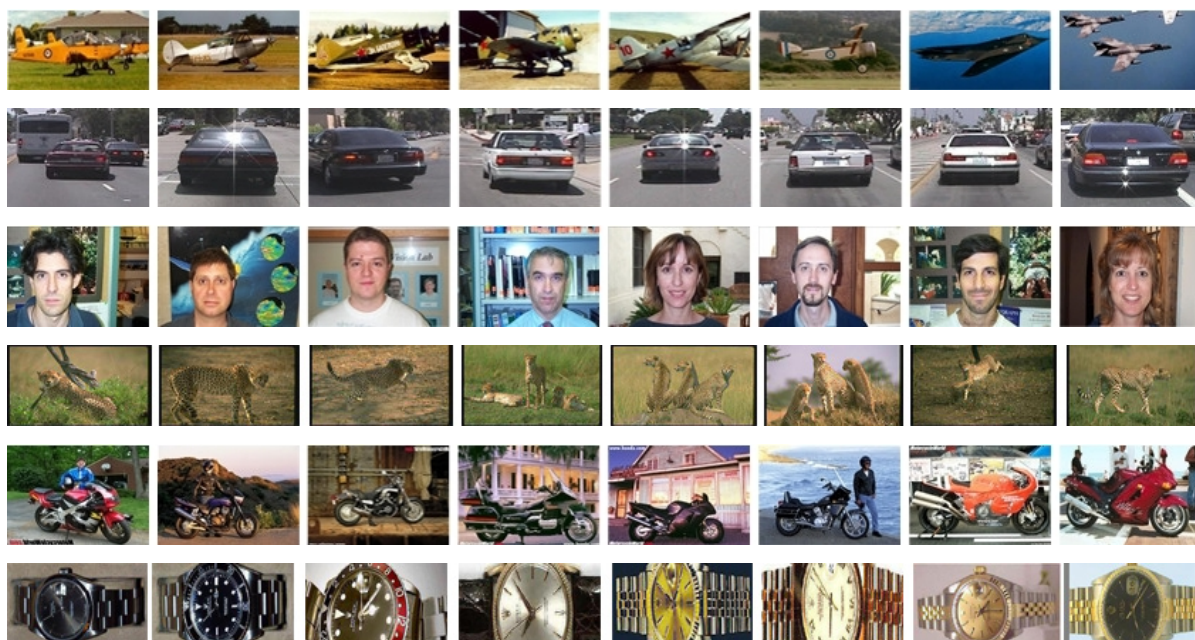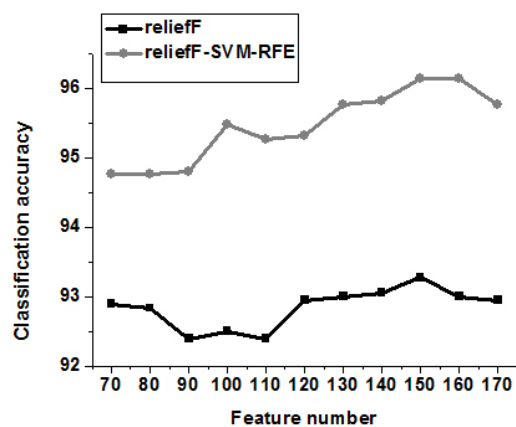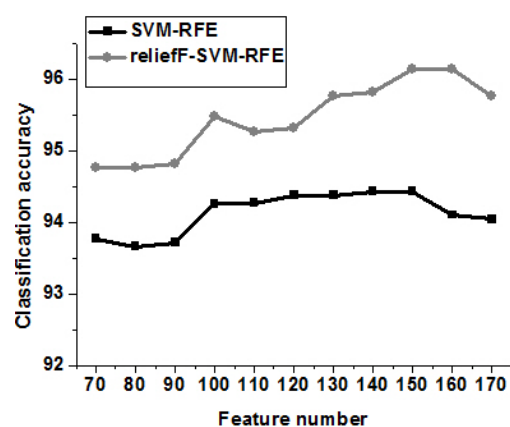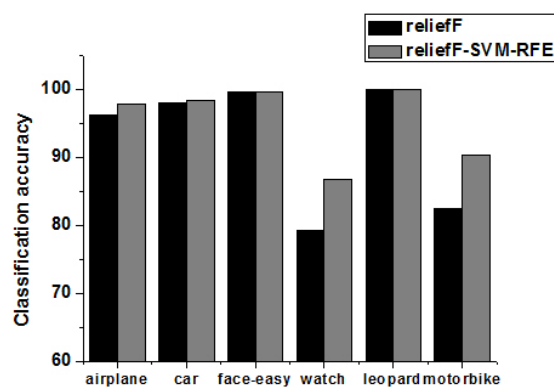
Figure 2. Some example images in dataset-Caltech.

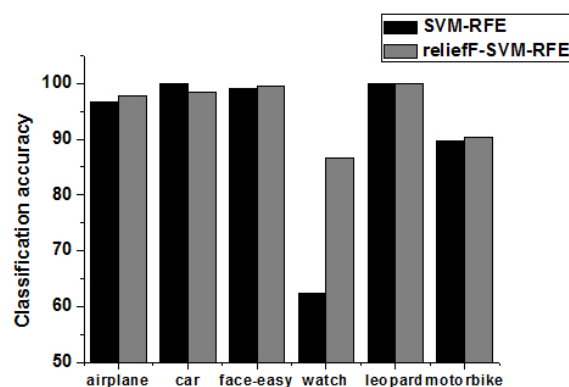**Figure 3.** Classification accuracy with different number of selected features. (a) and (b) Average accuracy for different number of features with reliefF, SVM-RFE and proposed algorithm. (c) and (d) Accuracy for each category based on the best average performance with reliefF, SVM-RFE and proposed algorithm.

Table 1. Computational time spent on each method.

Selection method	Computational time (Unit: s)
reliefF	143 s
SVM-RFE	35,752 s
reliefF-SVM-RFE	1715 s

The classification accuracy of each category achieved by three different algorithms are shown in **Figure 3(c)** and **Figure 3(d)**. It can be seen that our proposed algorithm performs the best among others on dataset-Caltech, which indicates the effectiveness of our new ranking criterion. When about 150 feature components are selected, our proposed reliefF-SVM-RFE method achieves the highest average classification accuracy of 96.14%, while the wrapper model SVM-RFE method achieves 94.43% and the filter model reliefF algorithm only achieves an accuracy of 93.27%. The results further demonstrate that our proposed hybrid model reliefF-SVM-RFE is superior to either the filter type relief method or the wrapper type SVM-RFE model.

Table 1 shows the computational time consumption of three algorithms. A Core 2 GHz personal computer running Matlab2009b is used in this experiment. As we can see that the filter model reliefF is the least computationally expensive while the wrapper model SVM-RFE is the most. 1715 seconds is required to select the final feature subset which means computational time spent on our proposed hybrid model is higher than reliefF but much lower than SVM-RFE. This result further proves that our model is not only effective but also efficient for feature selection in image classification.

5. Conclusion and Feature Work

In this paper, an improved and novel strategy combining reliefF and SVM-RFE algorithm is proposed to select a compact feature subset performing well in image classification. In this method, the reliefF method is employed to filter out many noisy feature components and obtain an effective candidate feature subset, then the new evaluation criterion integrating those from the reliefF method and the SVM-RFE is applied to select the final feature components. Experimental results demonstrate that with the feature subset from our proposed reliefF-SVM-RFE method a better classification performance can be achieved. However, some future work is needed to further improve the performance of the selection algorithm. For example, much more image descriptors can be extracted for images before processing or even a more efficient and less computational time consuming evaluation criterion can be devised. It is well recognized that the issue of effectively feature selection and accurately image classification still remains to be open.

Acknowledgements

References

[1] Guyon, I. and Elisseeff, A. (2003) An Introduction to Variable and Feature Selection. *The Journal of Machine Learning Research*, **3**, 1157-1182.

[2] Song, D.J. and Tao, D.C. (2010) Biologically Inspired Feature Manifold for Scene Classification, *IEEE Transaction on Image Processing*, **19**, 174-184.

[3] Marko, R.S. and Lgor, K. (2003) Theoretical and Empirical Analysis of ReliefF and RReliefF. *Machine Learning*, **53**, 23-69.

[4] Zhang, Y., Ding, C. and Li, T. (2008) Gene Selection Algorithm by Combining ReliefF and mRMR. *BMC Genomics*, 9, S27. http://dx.doi.org/10.1186/1471-2164-9-S2-S27

[5] Suykens, J.A.K. and Vandewalle, J. (1999) Least Squares Support Vector Machine Classifiers. *Neural Processing Letters*, **9**, 293-300. http://dx.doi.org/10.1023/A:1018628609742

[6] Guyon, I., Weston, J., Barnhill, S., *et al.* (2002) Gene Selection for Cancer Classification Using Support Vector Ma-

chines. *Machine Learning*, **46**, 389-422. http://dx.doi.org/10.1023/A:1012487302797

[7] John, G.H., Kohavi, R., Pfleger, K. (1994) Irrelevant Features and the Subset Selection Problem. *ICML*, **94**, 121-129.

[8] Yang, J., Yu, K., Gong, Y., *et al.* (2009) Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification. *IEEE Conference on Computer Vision and Pattern Recognition*, *CVPR* 2009, 1794-1801.

[9] Van De Sande, K.E.A., Gevers, T. and Snoek, C.G.M. (2010) Evaluating Color Descriptors for Object and Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**, 1582-1596. http://dx.doi.org/10.1109/TPAMI.2009.154

[10] Chun, Y.D., Kim, N.C. and Jang, I.H. (2008) Content-Based Image Retrieval Using Multiresolution Color and Texture Features. *IEEE Transactions on Multimedia*, **10**, 1073-1084. http://dx.doi.org/10.1109/TMM.2008.2001357

[11] Carlin, M. (2001) Measuring the Performance of Shape Similarity Retrieval Methods. *Computer Vision and Image Understanding*, **84**, 44-61. http://dx.doi.org/10.1006/cviu.2001.0935

[12] Mundra, P.A. and Rajapakse, J.C. (2010) SVM-RFE with MRMR Filter for Gene Selection. *IEEE Transactions on NanoBioscience*, **9**, 31-37. http://dx.doi.org/10.1109/TNB.2009.2035284