2010 3rd International Conference on Computer and Electrical Engineering (ICCEE 2010)
IPCSIT vol. 53 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V53.No.2.10

Feature Selection Method Based on Adaptive Relief Algorithm

Fan Wenbing+, Wang Quanquan and Zhu Hui

Zhengzhou University

Zhengzhou, China

Abstract—It is a very significant task that how to select informative features from the feature space for pattern recognition. Relief is considered one of the most successful algorithms for evaluating the quality of features. In this paper, we provide a valid proof for the first time, which demonstrates a blind selection problem in the previous Relief algorithm. Then we propose an adaptive Relief (A-Relief) algorithm to alleviate the deficiencies of Relief by dividing the instance set adaptively. Last, several experiments are reported by A-Relief and other feature selection methods. The experimental results illustrate the efficiency of A-Relief algorithm proposed in this paper.

Keywords-Relief algorithm; feature selection; pattern recognition;

1. Introduction

Feature selection is one of the important aspects of pattern recognition. A better feature selection algorithm, which eliminates the redundant feature effectively in feature space, can find a feature subset which is most relevant to models in current application. Not only can its proper design reduce system complexity, but it can also decrease processing time. Feature selection is widely used in image processing, feature reduction and machine learning as well as artificial intelligence, and it plays a critical role in many other cases. With limited training samples, selecting useful features for these kinds of problems poses a serious challenge to the existing feature selection algorithms.

Among the extant feature selection algorithms, the Relief algorithm is considered one of the most successful ones due to its simplicity and effectiveness. Relief algorithm was first proposed in [1]. The key idea of Relief is to iteratively estimate feature weights according to their ability to discriminate between neighboring models. Then, in [2] Relief was extended to handle noisy and missing data and solve multi-classification issues which the original Relief algorithm can not deal with. Subsequently, in order to explore the framework of expectation maximization, Iterative-Relief is put forward in [3]. Nevertheless, the deficiency of blind selection was not discovered in previous research.

This paper proposed a novel feature selection algorithm called Adaptive Relief (termed A-Relief). This proposed algorithm can divide the training set adaptively according to the peculiarity of these features. These features bring about blind selection when they are processing by former algorithms. Consequently, through handling these features by A-Relief, the authentic connection between features and models is reflected. It offers effective information for the further identification.

2. Scarcity of Relief Algorithm

Relief algorithm assesses the correlation between features and models by means of feature weight value, and yet in actual practice, there are still shortcomings in Relief algorithm, e.g., when all model types involved in the present problem are already definite, certain features still include a certain model type which is not

+ Corresponding author.
E-mail address: iewbfan@zzu.edu.cn

referred to in the current issue. In this case, these features, which are straightway substituted in Relief, are accounted to have an intimate relationship with model types, regardless of whether they related to model types. Therefore, Relief sometimes performs blind selection, which is not expected to occur. In the following, we provide a thorough interpretation of blind selection, which is never discovered in the anterior research.

The procedure of Relief algorithm is represented in Fig. 1 [1]. In each iteration, an instance x is randomly selected and then two nearest neighbors of x are found, one from the same classification (termed the *nearest hit* or NH) and the other from a dissimilar classification (termed the *nearest miss* or NM). The weight of the *i*th feature is then updated:

$$w_i = w_i + \left|\mathrm{x}^i - \mathrm{NM}^i(\mathrm{x})\right| - \left|\mathrm{x}^i - \mathrm{NH}^i(\mathrm{x})\right| \tag{1}$$

(1) Initialization: given $\Re = \{(\mathrm{x}_n, y_n)\}_{n=1}^{N}$, set $w_i = 0$, $1 \le i \le I$, number of iteration T;
(2) for $t = 1 : T$
  (3) Randomly select a instance x from $\Re$;
  (4) Find the nearest hit NH(x) and miss NM(x) of x;
  (5) for $i = 1 : I$
    (6) Compute: $w_i = w_i + \left|\mathrm{x}^i - \mathrm{NM}^i(\mathrm{x})\right| - \left|\mathrm{x}^i - \mathrm{NH}^i(\mathrm{x})\right|$
  (7) end
(8) end

Fig 1. Procedure of Relief algorithm

The Relief algorithm was incipiently designed to deal with binary problems. Afterward, Relief-F was proposed in [4] to dispose multiclass problems by perfecting the weight update rule (line 6 of Fig. 1) as:

$$w_i = w_i + \sum_{(c \in Y, c \neq y(x))} \frac{P(c)}{1 - P(c)} \{\left|x^i - \mathrm{NM}_c^i(x)\right| - \left|x^i - \mathrm{NH}_c^i(x)\right|\} \tag{2}$$

where $Y = \{1, \ldots, C\}$ is the model type space, $\mathrm{NM}_c(x)$ is the nearest miss of x from class c, and $P(c)$ is the a priori probability of class c.

From the above analysis, we find that Relief algorithm is a feature weighting algorithm that utilizes the performance of a highly nonlinear classifier in search for informative features.

In order to conveniently illustrate the major drawback of Relief, we provide two definitions as follows:

Definition 1: if feature η contains one or more model types, which model type space does not include in the problem to be resolved, we designate η as bogus feature.

Definition 2: a sort of model type, which does not exist in model type space in real application and is represented by bogus feature, is denoted as connotative classification (termed CC).

Compared with other features, bogus features usually perform some special particularities which can be summarized as:

- Regardless of whether there is strong correlation between bogus features and model type, iterated by Relief algorithm, the weights of bogus features achieve a larger value. Accordingly, the bogus feature is regarded as an informative feature which has a remarkable correlation with model type.
- Adopted a feature subset comprising the bogus feature, pattern recognition will deteriorate classification performance.

The description of bogus feature is presented in Fig 2. Fig 2 reveals the distribution of instance set of feature η. It is the ultimate purpose that the instance x can be accurately distinguished between model $class_A$ and model $class_B$. Consequently, model $class_C$ and model $class_D$ are unexpected model types which need not be transacted in this case (i.e., $class_C$ and $class_D$ are CC), and then feature η possesses the idiosyncrasy of bogus feature.

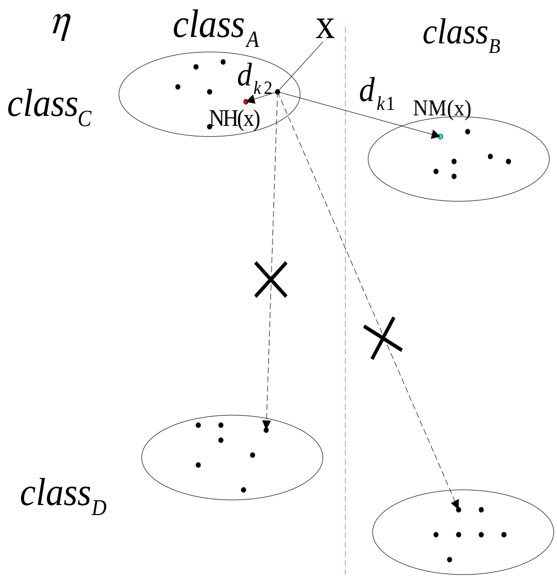

Fig 2. Distribution of feature η

We define the margin for a instance x as

$$\lambda_n = d(\mathrm{x}_n - \mathrm{NM}(\mathrm{x}_n)) - d(\mathrm{x}_n - \mathrm{NH}(\mathrm{x}_n)) \tag{3}$$

where NM(x) and NH(x) are the nearest miss and hit of instance x, respectively, and $d(\cdot)$ is the distance function. For a random instance x, if $x \in class_A$, then it is evident that $x \in class_C$ or $x \in class_D$; in this case, assuming $x \in class_C$, i.e., $x \in class_A \cap class_C$, we have $\lambda = d_{k1} - d_{k2} > 0$ .If $x \in class_B$, similarly we can conclude that d_{k1} is larger than d_{k2} , thus $\lambda > 0$. Hence, under the circumstance, for each instance x, $d(\mathrm{x}_n - \mathrm{NM}(\mathrm{x}_n)) - d(\mathrm{x}_n - \mathrm{NH}(\mathrm{x}_n)) > 0$ is established.

In the iteration procedure, substituted a positive value of λ_n into (1), the value of feature weight w_n keeps on increasing. Ultimately the feature η has been mistakenly considered as an informative feature due to the larger value of w_n, and obviously in Fig 2, there is few discrepancy between $class_A$ and $class_B$. Nevertheless, the bogus feature η, selected by Relief algorithm, participates in pattern recognition. It is reported that irrelevant features can deteriorate classification performance [5]. Therefore, blind selection in original Relief algorithm is a grievous mistake in the feature selection aspect, and then for classification purposes, removing irrelevant features is necessary.

3. Adaptive Relief Algorithm

Bogus features are unavoidable in several applications such as DNA microarray. Some researchers have indicated that the recognition of a small gene subclass with good predictive ability may not be sufficient to afford significant perspicacity into the understanding and modeling of the connection between certain diseases and genes [6].

In this section, A-Relief algorithm will be proposed to solve the blind selection problem in real application. Blind selection problems can not be easily settled through conventional optimization techniques. By compartmentalizing the instance subset adaptively according to the connotative classification presented in bogus features, the proposed algorithm implements an online algorithm that solves the blind selection problem. Accordingly, to discover bogus features is the major assignment.

Before giving the description of A-Relief algorithm we provide an interpretation of this algorithm as follows:

- According to the description of bogus features in Definition 1, the judgment fundament, which identifies bogus features, is whether the feature can contain the information of connotative classification.
- In the proposed algorithm, each feature will be inspected deeply to detect the bogus feature, before the feature is trained through A-Relief.

- If the present feature is a bogus feature, the instance subset will be divided by a threshold value ξ, which can be acquired by experiences or expert knowledge. For other features, they can be substituted into Relief algorithm straightway.

The procedure of A-Relief algorithm is presented in Fig. 3.

Prompt:
T number of iterations
I number of features
N number of instances (obviously $T \le N$)
Th a vector of threshold value
η_l the lth feature
$\Phi(\eta_l)$ the set of the values of feature η_l in all instances.
num a $I \times N$ matrix, the ith row of num (i.e., $num(i,:)$) is put into the index of a new subset, which is divided by Th .

(1) Initialization: given $\Re = \{(\mathrm{x}_n, y_n)\}_{n=1}^{N}$, set $w_i = 0$, $1 \le i \le I$;
(2) for $l = 1:I$
(3) Compute indexes: $num(l,:) = \{\Phi(\eta_l) > Th(l)\}$;
(4) Divide $\Re$ into different groups： $M_l = \{M_{l1},...,M_{lC}\}$
(5) end
(6) for $j = 1:I$
(7) if $\min\_dis(M_{j\cdot}, M_{j*}) > 3.5 \times \max\_dis(M_j)$
(8) Confirm: η_j is a bogus feature;
(9) for $k = 1:C$
(10) for $t = 1:T$
(11) Randomly select an instance x from M_{jk} ;
(12) Find the nearest hit $\mathrm{NH}(\mathrm{x})$ and miss $\mathrm{NM}(x)$ of x in M_{jk} ;
(13) Compute: $w_j = w_j + \left|\mathrm{x}^j - \mathrm{NM}^j(\mathrm{x})\right| - \left|\mathrm{x}^j - \mathrm{NH}^j(\mathrm{x})\right|$
(14) end
(15) end
(16) else for $t = 1:T$
(17) Randomly select an instance x from $\Re$;
(18) for $i = 1:I$
(19) Compute: $w_i = w_i + \left|\mathrm{x}^i - \mathrm{NM}^i(\mathrm{x})\right| - \left|\mathrm{x}^i - \mathrm{NH}^i(\mathrm{x})\right|$
(20) end
(21) end if
(22) end

Fig 3. Procedure of Adaptive Relief algorithm

The weight update rule (step 13 and 19 of Fig. 3) can be modified by (2) for the sake of handling multiclass problems. Obviously, compared with the previous Relief algorithm, the A-Relief algorithm adopts an approach that we differentiate the instances set by the connotative classification before training the feature. Then, for the bogus feature, the nearest hit $\mathrm{NH}(x)$ and miss $\mathrm{NM}(x)$ of x hail from not $\Re$ but M which can been obtained by Th (step 4 in Fig. 3).

$$M = \{class_C \cap class_A, class_C \cap class_B, class_C \cap class_A, class_C \cap class_B\} \tag{4}$$

In Fig. 2, based on the previous analysis, with trained by the original Relief algorithm, the feature η was considered one of optimization features. However, with the transformation of subset, which instance set was transformed from $\Re$ into M , η was regarded as an irrelevant feature owing to the diversification of training instance set by using A-Relief. This is in accord with the facts. In conclusion A-Relief successively performs online learning and solves the blind selection problems consisted in the original Relief algorithms.

4. Implementing and Analysis

In order to verify the effectiveness and efficiency of the Adaptive Relief for the blind selection problem, some images of train part are tested in this paper. It is our ultimate goal to identify fault images from all samples, which are shown in Fig. 4.

4.1. Calculating image features

We extract image features by the grey level co-occurrence matrix (termed GLCM), which is mentioned in [7]. The way of calculating the GLCM of image I is shown in Fig. 5, and there are four formats of GLCM, as showed in Fig. 6. For every format of GLCM, we can acquire six features as follows: *contrast*, *dissimilarity*, *homogeneity*, *entropy*, *energy*, *correlation* [8]. The abbreviation of *contrast0* in Table I means feature *contrast* is obtained through the 0^o direction GLCM in Fig. 6.

Consequently, we should compute 24 features aggregately in this case. In this image classification, we performed three sets experiments: 120 samples, 350 samples and 500 samples.

4.2. Results analysis

The Relief, Relief-F, I-Relief and A-Relief clustering approaches are applied on the three sets, respectively. Then, through a nonlinear classifier [9], features, selected by the three methods, were separately used to distinguish fault image from another set within 100 samples and 200 samples, in which samples were irrelevant to samples in the previous training sets.

According to the principle of the A-Relief algorithm in Fig.3, the weight of feature, iterated by the proposed algorithm, is the same as the weight trained by the original algorithms in addition to the weight of a bogus feature. Table I shows results of weights of feature *contrast*, which is inconsistent. Besides, feature *contrast* is the unique one in the feature space.

We can make a generalization that the feature *contrast* is a bogus feature. On further analysis, we detected that the feature *contrast* involved a connotative classification--exposure discrepancy. Owing to the complexity of shooting condition, it is ineluctable. There are abundant images exposed overly, as showed in the third picture of Fig. 4. Then by means of a nonlinear classifier, using the feature spaces chose respectively by the three methods, we evaluate the performance of each algorithm. Table II demonstrates the experimental results by using Relief, Relief-F, I-Relief and A-Relief. Through the comparison of the experiment in Table II, we can see that the ameliorative feature selection algorithm, A-Relief, improved the accuracy of the classification effectively. However, due to the presence of a bogus feature *contract* in feature space achieved by using Relief, Relief-F and I-Relief, their recognition results are extraordinary poor.

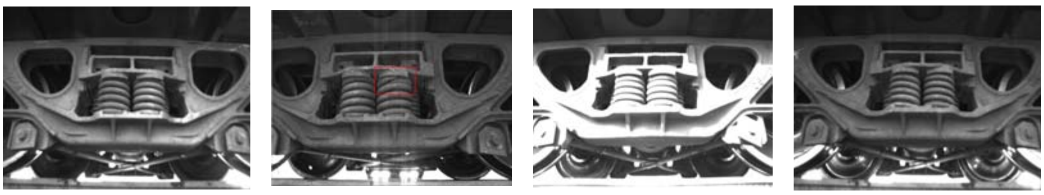

Fig 4. Some examples of train images

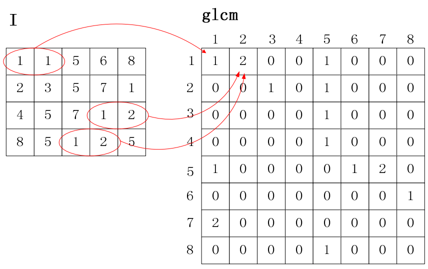

Fig.5. The way of calculating GLCM

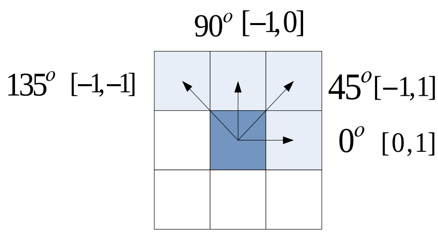

Fig.6. Four formats of GLCM

TABLE I. THE RESULT OF A BOGUS FEATURE TRAINED BY RELIEF, RELIEF-F, I-RELIEF AND A-RELIEF

Features	**Algorithm**	**Relief**			**Relief-F**			**I-Relief**			**A-Relief**		
	Sample number	120	350	500	120	350	500	120	350	500	120	350	500
Contrast0		0.978	0.965	0.982	0.896	0.868	0.882	0.764	0.782	0.731	0.125	0.130	0.129
Contrast45		0.904	0.915	0.912	0.825	0.831	0.826	0.826	0.797	0.819	0.033	0.039	0.036
Contrast90		0.942	0.921	0.948	0.873	0.861	0.874	0.678	0.691	0.706	0.062	0.067	0.068
Contrast135		0.928	0.920	0.919	0.842	0.836	0.848	0.724	0.763	0.701	0.039	0.041	0.034

TABLE II. CLASSIFICATION ACCURACY BY USING A-RELIEF AND THE PREVIOUS RELIEF ALGORITHM

Result	**Algorithm**	**Relief**		**Relief-F**		**I-Relief**		**A-Relief**	
	Sample number	100	200	100	200	100	200	100	200
right		26	47	60	118	75	142	92	186
wrong		74	153	40	82	25	58	8	14
Classification Accuracy		26%	23.5%	60%	59%	75%	71%	92%	93%

5. Conclusion

In this paper, we present an exhaustive interpretation of the blind selection problem existed in the original Relief algorithm, and the mathematical proof is provided. Then a novel feature selection has been proposed. We have adopted a technique based on differentiating the instances set adaptively in the proposed Adaptive Relief algorithm. Finally, we dealt with the train images by using the Adaptive Relief and the previous Relief algorithm. Experimental results illustrate that the amendatory Adaptive Relief algorithm improved the accuracy of the classification effectively and resolved the blind selection problem in the original algorithm drastically. How to decrease the complexity of this algorithm will be the further task of this paper.

6. References

[1] Kira K, Rendell L.A practical approach to feature selection[C].Proc 9th International Workshop on Machine Learning,1992:249-256.

[2] Robnik Sikonjam, Kononenko I.Theoretical and Empirical analysis of ReliefF and RReliefF. Machine Learning, 2003,53(1):23-69.

[3] Sun Yi Jun. Iterative relief for feature weighting algorithms, theories, and applications. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007,29(6):1035-1051.

[4] Kononenko I. Analysis and extensions of Relief. European Conference on Machine Learning,1994:171-182.

[5] R. Kohavi and G. H. John. Wrappers for Feature Subset Selection. Artifical Intelligence, vol. 97, nos. 1-2, pp. 273-324, 1997.

[6] T. Jenssen and E. Hovig, Gene-Expression Profiling in Breast Cancer. Lancet, vol. 365, pp. 634-635, 200.

[7] Gustavo Carneiro, Allan Jepson. Flexible spatial configuration of local image features. IEEE Transactions on Pattern Analysis and Machine Intelligence,2007,29(12):2089-2104.

[8] TATSUHIKO HIRUKAWA, SATOSHI KOMADA. Image feature based navigation of Nonholonomic Mobile Robots with active camera. SICE Annual Conference,2007:2502-2506.

[9] R. Duda, P. Hart, and D. Stork, Pattern Classification. J. Wiley, 2000.