Adaptive Relief Feature Evaluation and Selection based on Grey Level Co-occurrence Matrix

Han Wang[1], Zhousheng Ma[2], Wenbing Fan[1]
[1]School of Information Engineering, Zhengzhou University, Zhengzhou, Henan Province, China
[2]Henan Economic and Trade Vocational College, Zhengzhou, Henan Province, China
ernestwangxy@hotmail.com, 780697749@qq.com, iewbfan@zzu.edu.cn

Abstract - In image recognition, how to select informative features from the feature space is a very significant task. Relief algorithm is considered as one of the most successful methods for evaluating the quality of features. In this paper, it firstly provides a valid proof which demonstrates a blind selection problem in the previous Relief algorithm. And then this paper proposes an adaptive Relief (A-Relief) algorithm to alleviate the deficiencies of Relief by dividing the instance set adaptively. Lastly, it uses grey level co-occurrence matrix (GLCM) to extract text features and applies A-Relief algorithm to classify these features. The experimental results illustrate A-Relief algorithm proposed in this paper can improve the accuracy of the classification effectively and solve the blind selection problem.

Index Terms - adaptive Relief algorithm, feature selection, image recognition, GLCM.

1. Introduction

Feature selection is an important issue in pattern recognition and machine learning which helps us to focus the attention of a classification algorithm on those features that are the most relevant to predict the class. 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 problem. 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 discriminate between neighboring patterns. 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, with explored the framework of expectation maximization, Iterative-Relief is put forward in [3]. Nevertheless, the deficiency of blind selection was not discovered in previous research.

In this paper, a novel adaptive feature selection algorithm based on GLCM feature model is proposed. The algorithm is adaptive division of the training set 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. Finally, the experiments of the real train image targets are operated to indicate that the selections of feature generated by proposed algorithm have better performance.

2. Scarcity of Relief Algorithm

Relief algorithm embodies the correlation between features and models by means of the magnitude of feature weights, and yet in actual practice, there are still shortcomings in Relief algorithm. E.g., when all model types involved in the present problem have been definite, certain features still include model type which is not 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 are 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:

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

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

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

where $Y = \{1, ..., C\}$ 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 give two definitions as follows:

Definition 1: if feature η contains one or more model types that model type space do 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.

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

1) 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 the feature which has a remarkable correlation with model type.

2) Used a feature subset comprised the bogus feature, model identification 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, and then feature η possesses the idiosyncrasy of bogus feature.

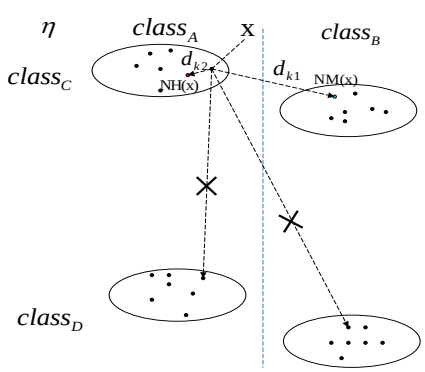

Fig. 2 Distribution of feature η

We define the margin for an instance x as

$$\lambda_n = d(x_n - \mathrm{NM}(x_n)) - d(x_n - \mathrm{NH}(x_n)), \quad (3)$$

where $\mathrm{NM}(x)$ and $\mathrm{NH}(x)$ 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 circumstances for each instance x, $d(x_n - \mathrm{NM}(x_n)) - d(x_n - \mathrm{NH}(x_n)) > 0$ is doubtless.

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 model 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 mantic 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, Adaptive Relief algorithm will be proposed to solve the blind selection problem in real application. Blind selection problems cannot be easily settled through conventional optimization techniques. With compartmentalized the instance subset adaptively by 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 Adaptive Relief algorithm we provide an interpretation of this algorithm as follows:

1) 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.

2) In the proposed algorithm, each feature will be inspected deeply to detect the bogus feature, before the feature is trained through Adaptive Relief.

3) 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 Adaptive 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 l th feature
$\Phi(\eta_l)$ the set of the values of feature η_l in all instances
num an $I \times N$ matrix, the i th row of num (i.e., $num(i,:)$) is deposited the index of a new subset, which is divided by Th

(1) Initialization: given $\Re = \{(x_n, y_n)\}_{n=1}^{N}$, 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 $\mathrm{NH}(x)$ and miss $\mathrm{NM}(x)$ of x in M_{jk} ;
        (13) Compute: $w_j = w_j + \left|x^j - \mathrm{NM}^j(x)\right| - \left|x^j - \mathrm{NH}^j(x)\right|$
      (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: $w_i = w_i + \left|x^i - \mathrm{NM}^i(x)\right| - \left|x^i - \mathrm{NH}^i(x)\right|$
      (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 Adaptive 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 M which can been implemented through Th in Fig. 2,

$$M = \{class_C \cap class_A, class_C \cap class_B, class_C \cap class_A, class_C \cap class_B\} \quad . (4)$$

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 instance set through the Adaptive Relief. This is in accord with the facts. In conclusion Adaptive 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 is shown in Fig.4.

A. *Calculating image features*

We extract image features by the grey level co-occurrence matrix (termed GLCM), which is mentioned in [7]. An approach for calculating the GLCM of image I is shown in Fig. 5, and there are four formats of GLCM, as shown 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 trained through the 0° 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.

B. *Results analysis*

The Relief, Relief-F, I-Relief and Adaptive clustering approaches are applied on the three sets, respectively. Then, through a nonlinear classifier [9], features, selected by the three methods, separately were 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 Adaptive 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 partial feature, the weight of 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 detect that the feature contrast is involved with a connotative classification--exposure discrepancy. Owing to the complexity of shooting condition, it is ineluctable. There are substantive images exposed overly, as shown 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 Adaptive Relief. Through the comparison of the experiment in Table II, we can see that the ameliorative feature selection algorithm, Adaptive Relief, improved the accuracy of the classification effectively. However, due to existing a bogus feature *contract* in feature space achieved by using Relief, Relief-F and I-Relief, their recognition results are extraordinary poor.

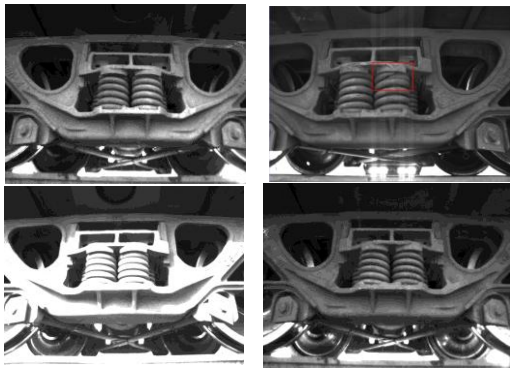

Fig. 4 Some examples of train images

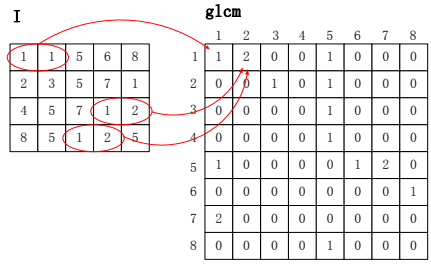

Fig. 5 The approach for calculating GLCM

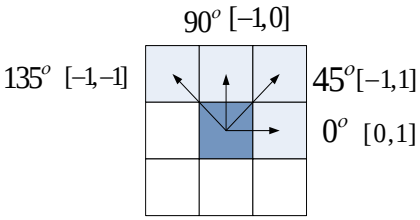

Fig. 6 Four formats for calculating GLCM

TABLE I The Weights of 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.

References

[1] Yuxuan Sun, Xiaojun Lou, Bisai Bao. A novel relief feature selection algorithm based on mean-variance model. Journal of Information & Computational Science, 2011, 8-16: 3921–3929

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

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

[4] Qinghua Hu, Xunjian Che, Lei Zhang, Daren Yu. Feature evaluation and selection based on neighborhood soft margin. Neurocomputing, 2010, 73: 2114 – 2124.

[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.