TELKOMNIKA, Vol. 11, No. 6, June 2013, pp. 3265 ~ 3270
e-ISSN: 2087-278X

Research of Plant-Leaves Classification Algorithm based on Supervised LLE

Yan Qing[1,2], Liang Dong*[2], Zhang Jingjing[1]
[1]School of Electrical Engineering and Automation, Anhui University, Hefei, Anhui, 230601, China
[2]School of Electronics and Information Engineering, Anhui University, Hefei, Anhui, 230601, China
*Corresponding author, e-mail: dliang@ahu.edu.cn

Abstract

A new supervised LLE method based on the fisher projection was proposed in this paper, and combined it with a new classification algorithm based on manifold learning to realize the recognition of the plant leaves. Firstly, the method utilizes the Fisher projection distance to replace the sample's geodesic distance, and a new supervised LLE algorithm is obtained .Then, a classification algorithm which uses the manifold reconstruction error to distinguish the sample classification directly is adopted. This algorithm can utilize the category information better, and improve recognition rate effectively. At the same time, it has the advantage of the easily parameter estimation. The experimental results based on the real-world plant leaf databases shows its average accuracy of recognition was up to 95.17%.

Keywords: ***r****ecognition of plant-leaves, supervised locally linear embedding, manifold distance, fisher projection*

1. Introduction

People generally select the local plant traits to identify plant such as leaf, flower, and fruit, stem or branch and so on. Due to collecting leaf is convenient and can maintaining plane form that suitable for two dimensions image procession, so leaf identification is the mostly direct and effective method to recognize plants.

The traditional method is extract classification characteristics of leaf image such as shape or texture traits, and then chose a classification algorithm to identify leaf [1-9]. But in real world, leaf images can be influenced easily by light, position, and size, so traditional feature extraction method is difficult to achieve accurately identification because of it can't adapt to changing environment. In recent years, researchers bring manifold methods into leaf identification and obtain some achievements [10-12]. Manifold learning as a new nonlinear dimension reduction method has become new research spot in data mining, pattern recognition and machine learning fields, and has applied to some fields successfully [13-16], such as face recognition and handwritting digital recognition.It aims to find low dimensional expressions of high dimensional manifold distribution data, and learns intrinsic geometric structure of low dimensional manifold from sampled data.

But the method still has trouble in low efficiency and accuracy [10] used ISOMAP manifold algorithm, its efficiency decline obviously with samples number increasing; [11] completed solution through the iteration, and also had the low efficiency problem. [12] applied the locally linear embedding algorithm, although it improved algorithm efficiency, identification accuracy still need to enhance.

Research finds that it exists two main elements to restrict identification accuracy: Firstly, LLE is a unsupervised algorithm essentially and can't apply samples' category information to classification and identification, so it influences the improving of identification accuracy. At present, some supervised LLE algorithms are proposed, but these methods are restrict learning sample space artificially, or add a penalty factor in heterogeneous sample neighbor distance to achieve supervision [17-20], they can't use category information of learning samples fully and need to chose many parameters. Secondly, the methods what is said above use manifold to extract feature and complete identification through classifier after reducing dimensions, but the

Received January 22, 2013; Revised April 9, 2013; Accepted April 23, 2013

choice of classifier can influence directly identification accuracy. Simple classifier is easy to achieve identify, but it has low accuracy; complex classifier can improve accuracy, but it will increase new estimating parameters probably which have large influence to recognition rate.

This paper has made the corresponding improvement according to two elements above. At first, it takes Fisher projection distance as geodesic distance to determine the neighborhood space in LLE. Because the distance is constructed to enhance sample divisibility, that is to say, close to intra-class samples and separate to interclass samples after reducing dimensions. Compare to traditional LLE method which uses Euclidean distance, Fisher projection distance can measure the distance differences between the samples better from the view of classification, mine samples' category information fully to improve the accuracy of identification and classification. Then, adopt a classification algorithm that use manifold to reconstruct error to complete identification of plant leaves [21]. This algorithm combines with manifold naturally and closely and needn't undertake new parameters selection at the same time, so it is easy to achieve and maintain stable high recognition rate.

2. Supervised LLE Algorithm based on Fisher Projection

Considering the shortage of traditional LLE that can't use supervised information, we adopt a new supervised LLE to achieve data dimensional reduction [22]. The procedure is following: Suppose data set $X\{x_1, x_2, ...x_n \in R^n\}$, every $\boldsymbol{x}_i$ is a linear combination with its K neighbor points.

(1) Project each sample point $\boldsymbol{x}_i$ to Fisher space using Fisher criterion, and then calculate projection distance between project point $\overline{x_i}$, and find K neighbors of $\boldsymbol{x}_i$ on the projection distance basis. Fisher Projection is committed to find the projection direction of the best classification rate from the view of classification, so the construction method of neighborhood is aim to classify and embody the category difference between the sample points better and is good to solve classify problems. However, traditional LLE constructs neighborhood by Euclidian distance. In the mapping process, Euclidian distance maps the points which far away in fact to neighbor points because of itself shortage, particularly in classify. The distance measure hardly embodies category feature, so it can't reflect the relationship between all kinds of samples and itself manifold.

(2) Reconstructed sample points by the neighborhood points on the basis of projection distance, and then calculate reconstruction weights W_{ij} which minimizes reconstruction cost function (5):

$$\varepsilon_X(W) = \sum_i \left| x_i - \sum_{j=1}^{K} W_{ij} x_j \right|^2 \quad (5)$$

In (5), when $\boldsymbol{x}_i$ is the neighbor of $\boldsymbol{x}_j$, W_{ij} satisfy the constraint $\sum_j W_{ij} = 1$ condition, otherwise, W_{ij} =0.

(3) Calculating low dimensional manifold on the basis of reconstruction coefficient matrix W. Minimize error function (6) to determine the coordinates of d dimensional manifold Y, that is to say, map $\boldsymbol{x}_i$ to d dimensional space $\boldsymbol{y}_i$.

$$\varepsilon_Y(Y) = \sum_{i=1}^{n} \left\| \boldsymbol{y}_i - \sum_{j=1}^{K} W_{ij} \boldsymbol{y}_j \right\|^2 \quad (6)$$

Where $\sum_i \boldsymbol{y}_i = 0$ and $\frac{1}{n}\sum_i \boldsymbol{y}_i \boldsymbol{y}_i^T = I$.

3. Manifold classification algorithm based on FS-LLE

Positive and negative sort of samples are in itself manifold respectively, as to a new sample, distance differences between it and positive and negative sort of samples manifold can be used as the basis for judging its category.FS-LLE reduces dimension through minimizing reconstruction error function $\varepsilon_Y(Y)$ in (6). It can be thought sample is closer to corresponding category manifold when the errors function smaller, so it can be used as a manifold distance measurement [18]. But LLE algorithm doesn't exist explicit mapping relationship from high dimension to low dimension, so it unable to incremental calculation. As to incremental sample, we can't calculate its project reconstruction error function because we can't find its low dimensional mapping. But LLE maintains the samples topological structure, [18] choses $\varepsilon_X(W)$ to replace $\varepsilon_Y(Y)$ as a manifold distance measurement to judge category.

The algorithm in this paper is constituted by FS-LLE and classification methods in [20].The algorithm can be divided into two parts that are learning procession and test procession. Specific steps are as follows:

(1) Learning procession:

(a) Projecting positive sample X^+ and negative sample X^- by Fisher criterion to Fisher space, and obtaining projection coordinate of each sample point. $\boldsymbol{X}^+ = \{\boldsymbol{x}_1^+, \boldsymbol{x}_2^+, ..., \boldsymbol{x}_m^+\}, \boldsymbol{X}^- = \{\boldsymbol{x}_1^-, \boldsymbol{x}_2^-, ..., \boldsymbol{x}_n^-\}$, and $\mathbf{x}_i^+, \mathbf{x}_i^- \in R^D$.

(b) As to positive sample X^+ and negative sample X^-, using FS-LLE to reduce their dimensions and obtaining low dimensional coordinate $Y^+ = \{\boldsymbol{y}_1^+, \boldsymbol{y}_2^+, ..., \boldsymbol{y}_m^+\}$ and $Y^- = \{\boldsymbol{y}_1^-, \boldsymbol{y}_2^-, ..., \boldsymbol{y}_n^-\}, \boldsymbol{y}_i^+, \boldsymbol{y}_i^- \in R^d$ and d<<D. D is the original sample dimension and d is the sample dimension after reducing dimension.

(2) Test procession

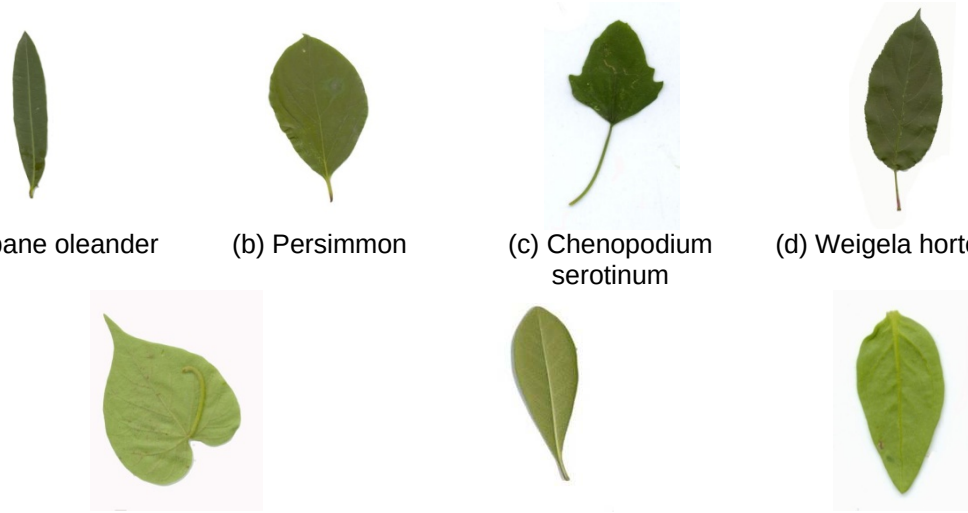
(a) As to test sample Z, projecting it to Fisher space and getting its projection coordinate. Then finding its K neighbors according to projection coordinate in positive sample X^+ and negative sample X^-. Adopting (5) to calculate positive and negative weight matrixes W^+, W^-.

Computing $\varepsilon_{X^+}(W^+)$ and $\varepsilon_{X^-}(W^-)$ respectively.

(b) Compared $\varepsilon_{X^+}(W^+)$ to $\varepsilon_{X^-}(W^-)$,if $\varepsilon_{X^+}(W^+) \geq \varepsilon_{X^-}(W^-)$,Z is belong to positive , otherwise, Z is belong to negative.

4. Classification Experiment

In order to verify the validity of the algorithm in this paper, we take experiments based on the plant leaf image database (http: / /www.intelengine.cn/data). This database which including 16846 leaf images belong to 220 species in total is provided by Hefei intelligence of the Chinese academic of sciences.The paper mainly discuses the leaf identification problem in two classes and designs two experiments. In the first experiment we select randomly 30 from 68 dogbane oleander images as positive sample, and 60 images from others species as negative sample; then selecting 60 images randomly from residual images (including dogbane oleander and others species) as test samples.In the second experiment we select randomly 20 samples respectively to construct train set from 6 kinds of leaf images, and select 10 images respectively randomly as test set from their residual images.The species include dogbane oleander, persimmon, chenopodium serotinum, weigela hortensis, ipomoea purpurea, and petunia. Parts of leaf images are showing in Figure 1. Choosing a species from train sample in turn in experiment, and other leaves as negative.



(a) Dogbane oleander (b) Persimmon (c) Chenopodium serotinum (d) Weigela hortensis

(e) Ipomoea purpurea (f) Pittosporum (g) Petunia

Figure 1. Positive and Negative Sort of Leaf Images

4.1. Preprocessing

The images in original database are not of uniform size and white background. In order to process conveniently, adjusting each image to 64×64 pixels and 256 levels gray image, then each leaf image can be expressed to a matrix and reshape the matrix to a row vector $I = \{i_{1,1}, ..., i_{1,64}, i_{2,1}, ..., i_{2,64}, ..., i_{64,64}\}$, so all samples can be taked as a 4096 dimensions vector.Considering the problem of small sample, we need to preprocess data set by PCA reducing dimensions before Fisher transformation. The experiment shows that when samples' dimensions down to 64 around can overcome the small sample problem.

4.2. Parameter Selection

In our experiment, neighbor parameter K and low dimension parameter d are the parameters that will influence classification results directly. In order to verify the effect of classification algorithm, experiment takes d for 2 uniformly.

As to selecting parameter K, we construct the sample's neighbors according the samples projection distance in Fisher space. The main idea of Fisher projection is separate categories extremely. But there are existing some samples which category features are different to the same class samples due to noise and nature conditions, the sample points named similar interference points have far distance with most of the same class points in Fisher projection distance. The neighbourhood may include more similar interference points possibly when K is too big. The presence of these points is adverse for accurate classification; otherwise neighbor structure can't reflect the manifold features of positive and negative category when K is too small. At first, the maximum neighbor parameter K is setted as 15 that is a half of the number of positive train samples, and its minimum is 5. In order to determine K, unchanging data set,we select the value of K from [5, 15] in turn, then adopt the manifold classification algorithm based on FS-LLE to test the recognition rate. Experiment results shows that the identification accuracy is changing in a small range when choose K from [6, 10]. The conclusion is similar to [20]. Considering the robustness of algorithm, we choose K as 7 in the first experiment.

4.3. Experiment result

In order to verify the effective of algorithm proposed in this paper, we have two group experiments. In the first, the data set is realized dimension reduction by FS-LLE, and then it is processed by the classification algorithm proposed in this paper、nearest neighbor algorithm and the SVM which kernel function is radial basis function. Each method carries on 20 times, the train samples and test samples are produced randomly from the corresponding sample set in every time. The average recognition rate and the best recognition rate in 20 experiments are shown in Table 1. Table 1 shows that the proposed algorithm improves identification accuracy

obviously. Although changing d in SVM can improve the identification accuracy, the effect is similar to the method proposed in this paper and it can increase the difficult in parameter selecting and algorithm complexity.

Table 1. Identification Accuracy in the First Experiment (%)

	Nearest neighbor	SVM	Classification algorithm in this paper
average recognition rate	91.77	93.03	95.17
the best recognition rate	93.33	96.67	98.33

The second experiment have 6 groups totally, each group chooses one from six species as positive samples. Each group repeats 10 times and calculates the average of the results, the value of K is 5 in the experiment. As to SVM, different kernel functions are chosen and the best recognition rate shows in Table 2. It is found that kernel function is important to SVM classifier, different kernel function can bring obviously different recognition result, but the classification results are stable by the algorithm proposed in this paper. From results, the recognition rates of persimmon. Weigela hortensis and petunia are low generally because of their similar shape; However, Chenopodium serotinum's recognition rate is high due to its different shape.

Table 2. Identification Accuracy in the Second Experiment (%)

Positive sample	Nearest neighbor	SVM (Polynomial kernel function)	SVM(radial basis function)	algorithm proposed in this paper
Persimmon	88.33	93.33	95	96.66
Chenopodium serotinum	90	95	96.66	96.66
Weigela hortensis	90	93.33	93.33	95
Ipomoea purpurea	91.67	93.33	95	96.66
pittosporum	93.33	95	96.66	96.66
Petunia	90	93.33	93.33	95

5. Conclusion

Traditional LLE relies on specific classifier to complete classification in low dimensional space when it is applied to identify leaves, so recognition results are related to classification algorithm closely; at the same time, the traditional LLE is restricted in classification and identification problem for it is a unsupervised method. In order to overcome the two shortages, it is proposed a method that constructed sample neighbourhood by Fisher projection distance, the improvement can utilize the category information of samples to reduce

Acknowledgements

This study was financed by the National Natural Science Foundation of China (61172127), the Research Fund for the Doctoral Program of Higher Education(KJQN1114), the 211 Project of Anhui University (KJQN1114). We also thank Dr. Abdul Wahid for his editing and improving of the paper.

References

[1] Du Ji-xiang, Wang Zeng-fu. Automatic identification method of plant leaf based on radial basis probabilistic neural network. *Pattern recognition and artificial intelligence*. 2008; 21(2): 206-212.
[2] Huang Lin, He Peng, Wang Jing-ming. Machine recognition research of plant leaf based on probabilistic neural network and fractal. *Journal of northwest agriculture and forestry university of science and technology (natural science edition)*. 2008; 36(9): 212-218.
[3] Su Yu-mei. Research and implementation of analysis method of plant leaf image. Master Thesis. Nanjing: Nanjing University of Science and Technology. 2007.

[4] Mokhtarian F, Abbasi S. Matching Shapes with Self-Intersection: Application to Leaf Classification. *IEEE Trans on Image Processing*. 2004; 13(5): 653-661.
[5] Oide M, Ninomiya S. Discrimination of Soybean Leaflet Shape by Neural Networks with Image Input. *Computers and Electronics in Agriculture*. 2000; 29(1): 59-72.
[6] Wang Xiaofeng, Huang Deshuang, Du Jixiang, et al. Feature extraction and recognition for leaf images. *Computer Engineering and Applications*. 2006; 42(3): 190-193.
[7] Zhang Lei. Research of plant species identification based on leaf features. Master Thesis. Changchun City: Northeast Normal University. 2007.
[8] Li Hui, Qi Lijun, Zhang Jian-hua, et al. Recognition of weed during cotton emergence based on principal component analysis and support vector machine. *Transactions of the Chinese society for agricultural machinery*. 2012; 43(9): 184-189.
[9] Li Xianfeng, Zhu Weixing, Kong Lingdong et al. Method of multi-feature fusion based on SVM and D-S evidengce theory in weed recognition. *Transactions of the Chinese society for agricultural machinery.* 2011; 42(11): 164-168.
[10] Zhang Shanwen, Huang De-Shuang. A robust supervised manifold learning algorithm and its application to plant leaf classification. *Pattern recognition and artificial intelligence*. 2010; 23(6): 836-841.
[11] Zhang Shanwen, Ju Chun-fei. Orthogonal global-locally discriminant projection for plant leaf classification. *Transactions of the Chinese Society of Agricultural Engineering*. 2010; 26(10): 162-166.
[12] Zhang Shanwen, Wang Xianfeng. Method of plant leaf recognition based on weighted locally linear embedding. *Transactions of the Chinese Society of Agricultural Engineering*. 2011; 27(12): 141-145.
[13] Xie Zhaoxia, Mu Zhichun, Xie Jian-jun. Multi-pose ear recognition based on locally linear embedding. *CAAI Transactions on Intelligent Systems*. 2008; 3(4): 321-327.
[14] Feng Hai-liang, LIi Jian-wei, Wang Xu-chu, et al. Face Recognition Based on Nonlinear Manifold Learning. *Journal of Chongqing University*. 2008; 31(3): 303-306.
[15] Jun Lai, Mei Xie. Segmentation Lung Fields in Thoracic CT Scans using Manifold Method. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(5): 1005-1014.
[16] Renbo Luo, Wenzhi Liao, Youguo Pi. Discriminative Supervised Neighborhood Preserving Embedding Feature Extraction for Hyperspectral-image Classification. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(5): 1051-1056.
[17] Chang Liu, JiLiu Zhou, Kun He, et al. Supervised Locally Linear Embedding in Tensor Space. *Third International Symposium on Intelligent Information Technology Application*. NanChang. 2009.
[18] Meng Deyu, Xu Zongben, Dai Ming Wei. A new supervised manifold learning method. *Journal of Computer Research and Development*. 2007; 44(12): 2072-2077.
[19] Kouropteva O, Okun O, Pietikainen M. *Supervised locally linear embedding algorithm for pattem recognition*. In Proc. IbPIuA 2003. LNCS 2652. Springer-Verlag. 2003:386-394.
[20] Ridder D de, Duin RPW. Locally linear embedding for classification. Technical Report PH-2002-01, Pattem Recognition Group, Dept of Imaging Science&Technology. Delft University of Technology, Delft. The Netherlands. 2002.
[21] Yao Li-Qun, Tao Qing. One Kind of Manifold Learning Method for Classification. *Pattern recognition and artificial intelligence*. 2005; 18(5): 542-545.
[22] Yan Qing, Liang Dong, Zhang Jingjing. Recognition method of plant leaves based on Fisher projection-supervised LLE algorithm. *Transactions of the Chinese society for agricultural machinery.* 2012; 43(9): 179-183.