2009 10th International Conference on Document Analysis and Recognition

Writer Adaptive Online Handwriting Recognition Using Incremental Linear Discriminant Analysis*

Zhibin Huang, Kai Ding, Lianwen Jin*, Xue Gao
College of Electronic and Information, South China University of Technology, Guangzhou, PRC
**E-mail: eelwjin@scut.edu.cn*

Abstract

Writer adaptive handwriting recognition, which has potential of increasing accuracies for a particular user, is the process of converting a writer-independent recognition system to a writer-dependent one. In this paper, we provide a general incremental learning solution for linear discriminant analysis (LDA) on the basis of previous researches, and propose an Incremental LDA (ILDA) based writer adaptive online handwriting recognition method. The adaptation is performed by modifying both the prototypes and the LDA transformation matrix through ILDA algorithm. It includes: (1) modifying prototypes in original feature space; (2) updating the LDA transformation matrix; (3) projecting the updated prototypes to LDA feature space. Experiments are performed on two datasets, the writer-dependent dataset, in which the writing style is consistent with the incremental training data, and the writer-independent dataset. The results demonstrated that our proposed method can reduce as much as 46.35% error rate on the writer-dependent dataset with only 0.20% accuracy loss on the writer-independent dataset. It indicates that our proposed method can significantly increase the recognition accuracy for a particular writer while has minor effects for general writers.

1. Introduction

The ability for a digital system to transcribe handwritten characters to a computerized text format is of great benefit in inputting, organizing and annotating data in various applications, such as the input, storage and distribution of notes or messages [1]. The success of products such as PDA and Tablet PC, is an evidence of the users' interest in such capabilities. On the other hand, recognition accuracy is the key factor in determining the acceptability of a handwriting recognition system. And some tests with keyboard typing have shown that the writer can tolerate random errors up to 1% while 0.5% is unnoticeable and 2% is intolerable [2]. However, due to various unconstrained cursive writing styles, the required accuracy is still too high to be satisfied for most of current handwriting systems.

It is generally agreed that, for a given handwriting recognition task, a writer-dependent (WD) system usually outperforms a writer-independent (WI) system. And writer adaption is the process of converting a WI system to a WD system. At present, a number of writer adaptation handwriting recognition methods have been proposed [3][4][5]. Vuori. V etc. [3] proposed a prototype based adaptation system using k nearest neighbor (KNN) classifier, the whole adaptation process includes three modules: adding new prototypes, deactivating confusing prototypes, and reshaping existing prototypes. Connel, SD and Jain, A.K. [4] proposed a adaptive online handwriting recognition model, where WI models is used to identify the styles present in a particular writer's training data, and then these models are retrained using the writer's data. Also, a self-growing probabilistic decision-based neural networks (SPDNNs) based adaptation method was proposed in [5].

On the other hand, since linear discriminant analysis (LDA) can find the linear projections of data which best separate two or more classes under the assumption that the classes have equal covariance Gaussian structure [6], it is widely employed in dimension reduction and feature extraction. This also motivates techniques of incremental linear discriminant analysis (ILDA) to deal with the situation that the complete set of training data is not all given in advance [7][8][9][10].

A number of researches about writer adaptation and ILDA were conducted. However, the ILDA based writer adaptation handwriting recognition remained unexploited. In this paper, we first provide a general incremental learning solution for LDA. Then an ILDA

* This paper is partially sponsored by the following research projects: NSFC (no. U0735004, 60772116), GDNSF (no. 07118074).

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.28

based writer adaptation handwriting recognition method is proposed. By implementing ILDA algorithm, the prototypes in the original feature space are modified first, then LDA transformation matrix is updated, finally the updated prototypes are projected to LDA space using updated LDA transformation matrix. The experimental results indicate that our proposed method can significantly reduce the error rate in the particular testing dataset while have limited effect on the general testing dataset at the same time.

The rest of this paper is organized as follows: Section 2 presents our general learning solution for the LDA algorithm, and then our adaptation method is proposed in Section 3. Section 4 describes the experiments and result. Finally, the conclusions are summarized in Section 5.

2. ILDA

Kim, T.K.et al. [9] proposed an incremental learning solution for LDA; however, the final solution of that method is too complex, since the sequential incremental learning condition and the chunk incremental learning condition are considered separately. And for each case, the solution is divided into two cases depending on whether a new class sample is added.

In this section, we first describe the principle of LDA. Then a general incremental learning solution for LDA is proposed based on the researches in [9].

2.1 LDA

Let us assume that ***N*** training samples ***X*** = *{x_i} (i=1,…,**N**)* in ***M*** classes have been presented so far. And $\boldsymbol{n_c}$ is the number of samples in class *c (c=1,…, **M**)*, such that $N = \sum_{c=1}^{M} n_c$, $\bar{x}_c$ and $\bar{x}$ are the mean vector of class *c* and all samples respectively.

According to [6], LDA seeks to find a linear transformation matrix W_{lda} over ***X*** in such a way that the ratio of the between-class scatter matrix S_b and the within-class scatter matrix S_w is maximized, where

$$S_b = \sum_{c=1}^{M} n_c(\bar{x}_c - \bar{x})(\bar{x}_c - \bar{x})^T \tag{1}$$

$$S_w = \sum_{c=1}^{M}\Sigma_c = \sum_{c=1}^{M}\sum_{j=1}^{n_c}(x_{cj} - \bar{x}_c)(x_{cj} - \bar{x}_c)^T \tag{2}$$

and x_{cj} strands for the *j*th sample in class *c*.

After that, the LDA transformation matrix W_{lda} can be obtained by conducting an eigenvalue decomposition of matrix $D = S_w^{-1} S_b$.

2.2 A General solution for ILDA

In derivation of ILDA, we assume that there are ***L*** incremental samples ***Y*** = *{y_i} (i=1,…,**L**) in **P** classes.* Without loss of generality, we assume that l_c of ***L*** incremental samples belong to class *c (c=1,…, **P**)*. Notice that, the class *c* may be the newly introduced class, in which case, $n_c = 0$ and $\bar{x}_c = 0$. Similarly with above assumption, the $\bar{y}_c$ and $\bar{y}$ represent the mean vector of class *c* and all incremental samples respectively. The within-class scatter matrix and between-class scatter matrix of the incremental samples are defined as S_{yw} and S_{yb} :

$$S_{yb} = \sum_{c=1}^{P} l_c(\bar{y}_c - \bar{y})(\bar{y}_c - \bar{y})^T \tag{3}$$

$$S_{yw} = \sum_{c=1}^{P}\Sigma_{yc} = \sum_{c=1}^{P}\sum_{j=1}^{l_c}(y_{cj} - \bar{y}_c)(y_{cj} - \bar{y}_c)^T \tag{4}$$

where y_{cj} strands for the *j*th sample in class *c*.

Since the new class may be introduced from incremental data and some classes may not be added any new samples, the merged class set ***Ω*** is divided into three parts: updated class set ***Ψ***, without updated class set ***Φ***, and newly introduced class set ***Γ***.

We assume the class number is updated to ***T (T* ≥ *M, T* ≥ *P)***, and the sample number of each class is $n'_c = n_c + l_c$, where *c=1,…,**T***. It is obvious that, if $c \in \Phi,\ l_c = 0$, and if $c \in \Gamma,\ n_c = 0$.

According to these assumptions, we can get that the updated mean vector of each class is

$$\overline{x'}_c = \frac{n_c \bar{x} + l_c \bar{y}}{n'_c} \ where\ c = 1,\ldots T \tag{5}$$

and the mean vector of total samples is

$$\overline{x'} = \frac{N\bar{x} + L\bar{y}}{N + L} \tag{6}$$

According to formulation (1) ~ (6), we can calculate the updated between-class scatter matrix S'_b and the within-class scatter matrix S'_w as follows:

$$S'_b = \sum_{c=1}^{T} n'_c(\overline{x'}_c - \overline{x'})(\overline{x'}_c - \overline{x'})^T \tag{7}$$

$$S'_w = \sum_{c=1}^{T}\Sigma'_c = \sum_{c=1}^{T}\sum_{j=1}^{n'_c}(x_{cj} - \overline{x'}_c)(x_{cj} - \overline{x'}_c)^T \tag{8}$$

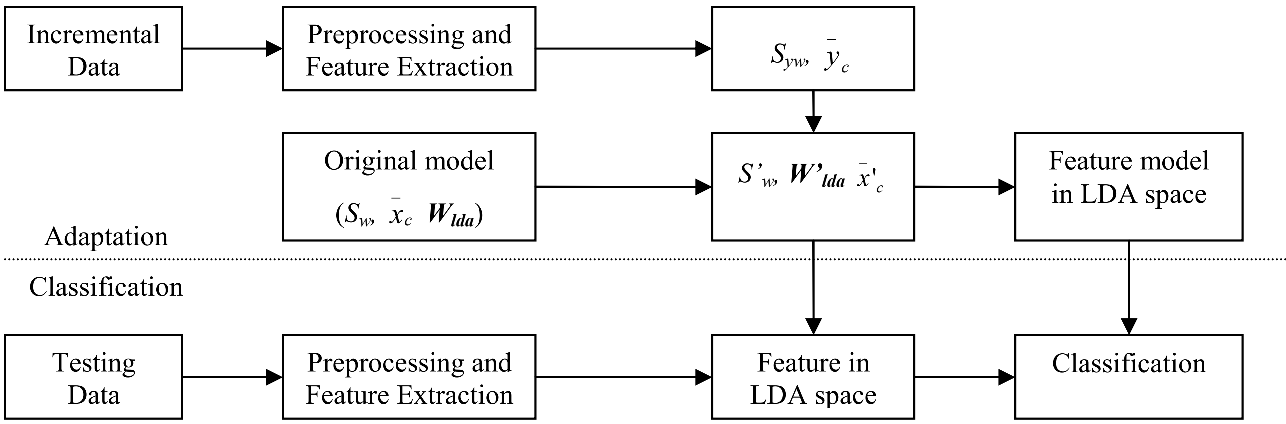

Figure1. **Diagram of the writer adaptive handwriting recognition system**

From [9], we can get that:

$$\Sigma'_c = \Sigma_c + \frac{n_c l_c^2}{(n_c+l_c)^2}(\bar{y}_c-\bar{x}_c)(\bar{y}_c-\bar{x}_c)^T + \frac{n_c^2}{(n_c+l_c)^2}\sum_{j=1}^{l_c}(y_{cj}-\bar{x}_c)(y_{cj}-\bar{x}_c)^T + \frac{l_c(l_c+2n_c)}{(n_c+l_c)^2}\sum_{j=1}^{l_c}(y_{cj}-\bar{y}_c)(y_{cj}-\bar{y}_c)^T \qquad (9)$$

For the last three terms

$$\begin{aligned}
&\frac{n_c l_c^2}{(n_c+l_c)^2}(\bar{y}_c-\bar{x}_c)(\bar{y}_c-\bar{x}_c)^T \\
&+\frac{n_c^2}{(n_c+l_c)^2}\sum_{j=1}^{l_c}(y_{cj}-\bar{x}_c)(y_{cj}-\bar{x}_c)^T \\
&+\frac{l_c(l_c+2n_c)}{(n_c+l_c)^2}\sum_{j=1}^{l_c}(y_{cj}-\bar{y}_c)(y_{cj}-\bar{y}_c)^T \\
&=\frac{1}{(n_c+l_c)^2}\{n_c^2[\sum_{j=1}^{l_c} y_{cj}y_{cj}^T - l_c\,\bar{y}_c\,\bar{y}_c^T] \\
&+[n_c l_c(l_c+n_c)(\bar{x}_c\,\bar{x}_c^T+\bar{y}_c\,\bar{y}_c^T-\bar{y}_c\,\bar{x}_c^T-\bar{x}_c\,\bar{y}_c^T)] \\
&+[(l_c^2+2n_c l_c)\sum_{j=1}^{l_c}(y_{cj}-\bar{y}_c)(y_{cj}-\bar{y}_c)^T]\} \\
&=\frac{1}{(n_c+l_c)^2}[(n_c+l_c)^2\Sigma_{yc} \\
&+n_c l_c(l_c+n_c)(\bar{y}_c-\bar{x}_c)(\bar{y}_c-\bar{x}_c)^T)] \\
&=\Sigma_{yc}+\frac{n_c l_c}{(n_c+l_c)}(\bar{y}_c-\bar{x}_c)(\bar{y}_c-\bar{x}_c)^T
\end{aligned} \qquad (10)$$

The within-class scatter matrix can be updated as:

$$\begin{aligned}
S'_w &= \sum_{c\in\Omega}\Sigma'_c = \sum_{c\in\Omega}\Sigma_c + \sum_{c\in\Omega}\Sigma_{yc} \\
&+\sum_{c\in\Omega}\frac{n_c l_c}{n_c+l_c}(\bar{y}_c-\bar{x}_c)(\bar{y}_c-\bar{x}_c)^T \\
&= \sum_{c\in(\Omega-\Gamma)}\Sigma_c + \sum_{c\in\Gamma}\Sigma_c + \sum_{c\in(\Omega-\Phi)}\Sigma_{yc} + \sum_{c\in\Phi}\Sigma_{yc} \\
&+\sum_{c\in\Psi}\frac{n_c l_c}{n_c+l_c}(\bar{y}_c-\bar{x}_c)(\bar{y}_c-\bar{x}_c)^T \\
&+\sum_{c\in(\Omega-\Psi)}\frac{n_c l_c}{n_c+l_c}(\bar{y}_c-\bar{x}_c)(\bar{y}_c-\bar{x}_c)^T
\end{aligned} \qquad (11)$$

It is obvious that:

$$\begin{aligned}
&\sum_{c\in(\Omega-\Psi)}\frac{n_c l_c}{n_c+l_c}(\bar{y}_c-\bar{x}_c)(\bar{y}_c-\bar{x}_c)^T \\
&=\sum_{c\in\Gamma}\Sigma_c = \sum_{c\in\Phi}\Sigma_{yc} = 0
\end{aligned} \qquad (12)$$

Insert formulation (2) (4) (12) to (11), we can get:

$$S'_w = S_w + S_{yw} + \sum_{c\in\Psi}\frac{n_c l_c}{n_c+l_c}(\bar{y}_c-\bar{x}_c)(\bar{y}_c-\bar{x}_c)^T \qquad (13)$$

According to formulation (7) and (13), we can calculate the matrix $D' = S'^{-1}_w S'_b$, and then obtain the updated LDA transformation matrix W'_{lda} by conducting an eigenvalue decomposition on D'.

3. Adaptation

The diagram of our proposed adaptive method is shown in Figure1. The whole adaptation procedure can be described in the following four steps:

(1) Extract the feature of incremental data, in this paper, 8-directional feature proposed by ZL. Bai and Q. Huo[11] is employed, the original feature dimension being 512.

(2) Calculate the within-class scatter matrix and the mean vector of each class for incremental data.
(3) Update the within-class scatter matrix, mean vector of each class and the LDA transformation matrix by implementing ILDA algorithm.
(4) Project the mean vector of each class to the LDA feature space and reduce the feature dimension to 160 by multiplying the updated LDA matrix W'_{lda}.

For classification, we first extract the 8-directional feature of the testing data, and then project the original feature to the LDA feature space and reduce the feature dimension through updated LDA matrix W'_{lda}. Finally, the features of testing data and the model are matched in the LDA feature space and output the recognition result.

4. Experiments and results

4.1 Experimental data

The benchmark data used in this paper comes from the SCUT-COUCH database. It is a revision of SCUT-COUCH2008 [12], which is now contributed by more than 168 participants. All characters are written in an unconstrained manner. This database is a comprehensive dataset composed of 8 subsets: GB1 (level 1 GB2312-80) simple Chinese character, GB2 (level 2 GB2312-80) simple Chinese character, traditional Chinese character, word, Pinyin, digit, alphabet and symbol. (The SCUT-COUCH database is available at: http://www.hcii-lab.net/data/SCUTCOUCH/).

In this paper two subsets of SCUT-COUCH dataset are used. One is the GB1 subset which contain 168 writers' sample of **3755** categories of simple Chinese characters. The other is the Word8888 subset, which consists of 30 writers' samples of **8888** categories of word subset. All of the word data are manually segmented into isolated characters, which contain 2078 categories of 19595 isolated GB1 simple Chinese characters. In other words, we have a writer-independent dataset that contains 168 sets of 3755 classes of GB1 Chinese characters, and a writer-dependent dataset that contains 30 sets of 2078 classes of GB1 Chinese characters. The writer-independent dataset will be used to train a baseline classifier, and the writer-dependent dataset will be used to train/test the ILDA model for writer adaptation.

4.2 Experimental setup

In the following experiments, we randomly select 134(or 79.16%) sets of data from the writer-independent dataset for training. And the remaining 34(or 20.84%) sets are treated as the writer-independent testing dataset to evaluate the effect of our proposed adaptive method for general writers. For each particular writer's handwritting samples, which are obtained from the writer-independent dataset. We randomly select 50% of each category's data for learning the ILDA model, and then use the remaining data to test the writer adaption performance.

From the formulation (1) ~ (13), it is obvious that the performance of adaptation is relative to the number of updating samples. However, since the frequencies of the characters which can be found in the word subset are different from each other, the updating sample number of each category is not the same. To overcome this problem, the updating number of each category is normalized to a fixed value: $l_c' = r \times n_c \quad c = 1,\dots,T$, where r is the parameter. According to this, the formulation (4) ~ (6) are modified as follows:

$$S_{yw} = \sum_{c=1}^{P} \Sigma_{yc} = \sum_{c=1}^{P} \frac{r \times n_c}{l_c} \sum_{j=1}^{l_c} (y_{cj} - \bar{y}_c)(y_{cj} - \bar{y}_c)^T \quad (14)$$

$$\bar{x}'_c = \frac{n_c \bar{x}_c + (r \times n_c)\bar{y}_c}{(1+r)n_c} = \frac{\bar{x}_c + r \times \bar{y}_c}{(1+r)} \; where \; c = 1,\dots,T \quad (15)$$

$$\bar{x}' = \frac{N\bar{x} + (N \times r)\bar{y}}{(1+r) \times N} = \frac{\bar{x} + r \times \bar{y}}{(1+r)} \quad (16)$$

4.3 Performance on writer-dependent dataset

The first experiment is designed to compare the performance of using adaptation and not-using adaptation on the writer-dependent testing dataset. And the recognition accuracy is given in Table1.

Table1. **Performance comparison on writer-dependent dataset**

r	Without adaptation	With adaptation	Error rate reduction
0.05	83.02%	84.68%	9.81%
0.1		86.49%	20.44%
0.2		89.43%	37.75%
0.3		90.89%	46.35%
0.4		91.96%	52.65%
0.5		92.39%	55.18%
0.6		92.82%	57.71%

From Table1, it is observed that our writer adaptation method can significantly reduce the error rate for the 30 sets of writer-dependent dataset, and the recognition accuracy is increased with the increase of the updating ratio *r*.

4.4 Performance on writer-independent dataset

Besides evaluating the performance on writer-dependent testing dataset, another experiment is performed to evaluate the effect of the proposed adaptation method on writer-independent testing dataset. The result is shown in the Table2.

As shown in Table2, although the recognition accuracy is decreased after adaptation, the accuracy loss is very small, especially for small values of *r*. This indicates that while the writer adaption can significantly reduce the error rate for writer-dependent dataset, it has limited negative effect on a writer-independent testing dataset.

Table2. **Performance comparison on writer-independent dataset**

r	Without adaptation	With adaptation	Accuracy loss
0.05	93.83%	93.63%	0.2%
0.1		93.58%	0.25%
0.2		93.32%	0.51%
0.3		92.98%	0.85%
0.4		92.50%	1.33%
0.5		91.96%	1.87%
0.6		91.33%	2.5%

From Table1 and Table2, it is obvious that when $r < 0.1$, the error rate reduction on writer-dependent dataset is much smaller; and when $r \geq 0.4$, the proposed method may lose more than 1% accuracy on writer-dependent testing dataset. Therefore, we suggest that the reasonable range of *r* be from 0.1 to 0.3. Under these conditions, our proposed adaptation method can reduce about 20.44%~46.35% error rate on the particular writer-dependent dataset while only has less than 0.85% accuracy loss on the writer-independent dataset.

5. Conclusion

In this paper, we first present a general solution for ILDA, and then propose an ILDA based writer adaptive handwriting recognition method. Experiments are performed on two datasets, one being the writer-dependent dataset, in which the writing style is consistent with the incremental training data, and the other a writer-independent dataset. From these experiments, it can be found that our proposed method can significantly reduce the error rate for the particular writers while have limited effects for general writers.

6. References

[1] Connel, S.D. and Jain, A.K., "Writer Adaptation of Online Handwriting Models", *ICDAR99*, 1999, pp.434—437.

[2] Cole, R.A. and Mariani, J. etc. *Survey of the State of the Art in Human Language Technology*, Cambridge University Press Cambridge, UK, 1997.

[3] Vuori, V. and Teknillinen Korkeakoulu, *Adaptive methods for on-line recognition of isolated handwritten characters*, Helsinki University of Technology, Helsinki, 2002

[4] Connel, SD and Jain, A.K., "Writer Adaptation of Online Handwriting Models", *IEEE Transaction on PAMI, Vol. 24, No. 3*, 2002, pp. 329-346.

[5] Fu, H.C. and Chang, H.Y. etc., "User Adaptive Handwriting Recognition by Self-Growing Probabilistic Decision-Based Neural Networks", *IEEE Transaction on Neural Networks*, *Vol. 11, No. 6*, 2000, pp. 1373-1384

[6] Duda, R.O. and Hart, P.E. and Stork, D.G. *Pattern Classification*. Wiley New York, New York, 2001.

[7] Jing, X.Y. and Zhang, D. and Tang, Y.Y., "An Improved LDA Approach", *IEEE Transactions on SMC, Vol. 34, No. 5*, 2004, pp. 1942-1951

[8] Zhang, P. and Peng, J. and Riedel, N., "Discriminant Analysis: A Least Squares Approximation View", *CVPR2005*, 2005, pp. 46-53

[9] Pang, S. and Ozawa, S. and Kasabov, N., "Incremental linear discriminant analysis for classification of data streams", *IEEE Transactions on SMC, Vol. 35, No. 5*, 2005, pp. 905-914

[10] Kim, T.K. and Wong, S.F. et al., "Incremental Linear Discriminant Analysis Using Sufficient Spanning Set Approximations", *CVPR2007*, 2007, pp. 1-7

[11] Zhen-Long, BAI and Qiang, HUO, "A Study On the Use of 8-Directional Features For Online Handwritten Chinese Character Recognition", *ICDAR2005*, 2005, pp. 232-236

[12] Yunyang Li, Lianwen Jin, Xinghua Zhu, Teng Long, "SCUT-COUCH2008: A Comprehensive Online Unconstrained Chinese Handwriting Dataset", *ICFHR2008*, 2008, pp. 165-170.