Journal of Southeast University (English Edition) Vol. 32, No. 1, pp. 35 - 38 Mar. 2016 ISSN 1003—7985

Enhanced kernel minimum squared error algorithm and its application in face recognition

Zhao Yingnan[1,2,3] He Xiangjian[3] Chen Beijing[1,2] Zhao Xiaoping[1,2]

([1] Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing 210044, China)

([2] School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China)

([3] School of Computing and Communications, University of Technology, Sydney, Sydney 2007, Australia)

Abstract: To improve the classification performance of the kernel minimum squared error (KMSE), an enhanced KMSE algorithm (EKMSE) is proposed. It redefines the regular objective function by introducing a novel class label definition, and the relative class label matrix can be adaptively adjusted to the kernel matrix. Compared with the common methods, the new objective function can enlarge the distance between different classes, which therefore yields better recognition rates. In addition, an iteration parameter searching technique is adopted to improve the computational efficiency. The extensive experiments on FERET and GT face databases illustrate the feasibility and efficiency of the proposed EKMSE. It outperforms the original MSE, KMSE, some KMSE improvement methods, and even the sparse representation-based techniques in face recognition, such as collaborate representation classification (CRC).

Key words: minimum squared error; kernel minimum squared error; pattern recognition; face recognition

DOI: 10.3969/j.issn.1003 - 7985.2016.01.007

The kernel method is one of the most influential developments in recent decades, and many kernel-based techniques have been developed. Among these methods, the kernel minimum squared error (KMSE) has gained much attention due to its high computational efficiency[1]. Recent study has illustrated the good generalization power of KMSE. It is a unified framework[2] for some known algorithms, such as KFDA, least squares SVM (LS-SVM), and the kernel ridge regression (KRR). Meanwhile, KMSE has been applied in the sparse representation-based methods[3-4]. Similar to other kernel methods, KMSE suffers from a problem during the training phase, that is, the computation efficiency of KMSE decreases as the number of the training patterns increases. Another drawback lies in the uncertainty of KMSE solutions, since the number of coefficients is greater than that of linear equations. To tackle these problems, KMSE can be amended through either accelerating or model improvement.

Most of existing KMSE improvements focus on addressing the first problem. Generally, these methods select so-called significant nodes (SNs) from the training samples, which have great contributions to classification according to some rules. The number of SNs is always smaller than that of training ones. Hence, in the classification phase, these SNs substitute all the training samples to compute the kernel functions. Accordingly, they can speed up the algorithms while not sacrificing the classification rates. Simplified KMSE (SKMSE)[5] applies the distance criterion to select SNs. More recently, Zhao et al.[6] presented a backward greedy searching strategy. In addition, Zhao et al.[7] employed the incremental learning technique to speed up the KMSE algorithm.

Except for the accelerating algorithms, some literature enhances KMSE for efficient classification in terms of model improvement. Taking advantage of the local ridge regression (LRR), Xu et al.[8] developed an efficient KMSE classification approach. It appears complex although it adopted the iterative technique for simplification. Wang[9] presented the minimum norm minimum squared-error (MNMSE) algorithm to improve the classification and numerical stability of KMSE. In this algorithm, the definition of geometric distance between classes, which is evaluated by $\boldsymbol{w}$ vector of the KMSE model, is introduced to modify the objection function.

We modify the KMSE objective function by adding an adaptive item called enhanced KMSE (EKMSE). This algorithm enables different classes to have greater difference to achieve better classification performance.

1 KMSE Overview

Suppose that there are c classes and each class has N sam-

Received 2015-10-07.

Biography: Zhao Yingnan (1973—), female, doctor, associate professor, ann_zhao_99@163.com.

Foundation items: The Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), the National Natural Science Foundation of China (No. 61572258, 61103141, 51405241), the Natural Science Foundation of Jiangsu Province (No. BK20151530), Overseas Training Programs for Outstanding Young Scholars of Universities in Jiangsu Province.

Citation: Zhao Yingnan, He Xiangjian, Chen Beijing, et al. Enhanced kernel minimum squared error algorithm and its application in face recognition[J]. Journal of Southeast University (English Edition), 2016, 32 (1): 35 - 38. DOI: 10.3969/j.issn.1003 - 7985.2016.01.007.

ples. We consider the input-output data model of KMSE as

$$\boldsymbol{G} = \boldsymbol{\phi W} + \boldsymbol{E} \tag{1}$$

where

$$\boldsymbol{\phi} = \begin{bmatrix} 1 & \boldsymbol{\phi}(x_1)^{\mathrm{T}} \\ \vdots & \vdots \\ 1 & \boldsymbol{\phi}(x_l)^{\mathrm{T}} \end{bmatrix},\ \boldsymbol{W} = \begin{bmatrix} w_0 \\ \boldsymbol{w} \end{bmatrix},\ \boldsymbol{E} = \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix},\ \boldsymbol{G} = \begin{bmatrix} g_1 \\ \vdots \\ g_N \end{bmatrix}$$

and $\boldsymbol{W}$ is the column vector containing the bias item w_0 and the normal vector $\boldsymbol{w}$. Here, $\boldsymbol{\phi}_i(x)$ denotes the i-th feature of sample x in kernel space. According to the reproducing kernel theory, there exist the coefficients $\alpha_i = 1, 2, \ldots, N$ satisfying

$$\boldsymbol{W} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\phi}(x_i) \tag{2}$$

Substituting Eq. (2) into Eq. (1), we obtain

$$\boldsymbol{KA} + \boldsymbol{E} = \boldsymbol{G}, \quad \boldsymbol{A} = \{w_0, \alpha_1, \alpha_2, \ldots, \alpha_l\}^{\mathrm{T}} \tag{3}$$

where $\boldsymbol{K} = \begin{bmatrix} 1 & k(x_1, x_1) & \ldots & k(x_1, x_N) \\ 1 & k(x_2, x_1) & \ldots & k(x_2, x_N) \\ \vdots & \vdots & & \vdots \\ 1 & k(x_N, x_1) & \ldots & k(x_N, x_N) \end{bmatrix}$. $\boldsymbol{K}$ is referred to as the kernel matrix of KMSE. Eq. (3) can be solved using the least squared criterion. We have

$$\boldsymbol{A} = (\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K} + \gamma \boldsymbol{I})^{-1}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{G} \tag{4}$$

For an arbitrary sample x, we calculate the distances between g_x and the class labels of all the c classes, and then classify it to the closest one. The equation is

$$\boldsymbol{g}_x = w_0 + \boldsymbol{k}_x \boldsymbol{\alpha} \tag{5}$$

where $\boldsymbol{k}_x = [k(x, x_1), \ldots, k(x, x_N)]$.

2 Enhanced Kernel Minimum Squared Error (EKMSE) Classification

On the basic idea of KMSE, we modify the class label definition by adding an adaptive item to enlarge the distance of different classes. At the same time, we adopt an iteration technique to determine the two corresponding parameters.

Eq. (1) is modified as

$$\boldsymbol{G} + \boldsymbol{\mu S} = \boldsymbol{\phi W} + \boldsymbol{E} \tag{6}$$

where $\boldsymbol{S}$ is the variation of the class label $\boldsymbol{G}$ by substitute 0 with -1. Such definition will enlarge the class difference. The items in $\boldsymbol{G}$ are 0 and 1 and they are -1 and 1 in $\boldsymbol{S}$. $\boldsymbol{\mu}$ is the self-adapting item, varying with $\boldsymbol{K}$, $\boldsymbol{A}$ and $\boldsymbol{G}$. Therefore, we can obtain the objective function of EKMSE as

$$\boldsymbol{KA} + \boldsymbol{E} = \boldsymbol{G} + \boldsymbol{\mu S} \tag{7}$$

By virtue of the least squared criterion, we obtain

$$\boldsymbol{A} = (\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K} + \lambda \boldsymbol{I})^{-1}\boldsymbol{K}^{\mathrm{T}}(\boldsymbol{G} + \boldsymbol{\mu S}) \tag{8}$$

Meantime, we obtain the following on the base of Eq. (7):

$$\boldsymbol{\mu S} = \boldsymbol{KA} - \boldsymbol{G} \tag{9}$$

Given a small positive initial value of $\boldsymbol{\mu S}$, $\boldsymbol{\mu S}$ and $\boldsymbol{A}$ are calculated iteratively.

The algorithm can be summarized in the following steps:

Step 1 Define the class label matrix $\boldsymbol{G} = \{g_i\}_{i=1}^{N}$ of the training samples. Suppose that there are N training samples and c classes.

Step 2 Define matrix $\boldsymbol{S}$ on the base of $\boldsymbol{G}$, substituting 0 with -1, the left numbers in $\boldsymbol{G}$ are unchanged.

Step 3 Calculate $\boldsymbol{K}$ with the kernel function.

Step 4 Calculate $\boldsymbol{A}$ and $\boldsymbol{\mu S}$ using the iterative method.

First, initialize an $N \times c$ matrix $\boldsymbol{\mu}_0$ randomly. Then point-divide $\boldsymbol{\mu}_0$ with a large number M, and point-multiples matrix $\boldsymbol{S}$ finally, i.e.,

$$\boldsymbol{\mu S} = \boldsymbol{\mu}_0 \boldsymbol{S}/M \tag{10}$$

Then, calculate $\boldsymbol{A}$ and $\boldsymbol{\mu S}$ iteratively using Eq. (8) and Eq. (9). Note that we assign 0 for each negative element in $\boldsymbol{\mu S}$ to guarantee its non-negative characteristic.

Step 5 Calculate the system output $\tilde{\boldsymbol{g}}_x$ between the test sample x and the training ones using the kernel function again:

$$\tilde{\boldsymbol{g}}_x = \left[\sum_j k(x, x_j)\alpha_{j1}, \ldots, \sum_j k(x, x_j)\alpha_{jc}\right] = \boldsymbol{T}_x \boldsymbol{A} \tag{11}$$

Step 6 Let $d_i = \| \boldsymbol{g}_i - \tilde{\boldsymbol{g}}_x \|_2^2$ stand for the distance between the test sample and i-th class. If $r = \arg\min_j d_j$, then the test sample r is assigned to the r-th class.

To analyze the difference between EKMSE and KMSE, we give the definition of predicted error as $\| \boldsymbol{C}_t - t\overline{\boldsymbol{G}} \|$, where C_t stands for the true class label of sample t, and $\overline{G}$ is the classification result of KMSE or EKMSE. Obviously, the smaller the predicted error is, the better performance of the algorithm to approximate the true sample. Fig. 1 implies that EKMSE usually has much fewer predicted errors than KMSE.

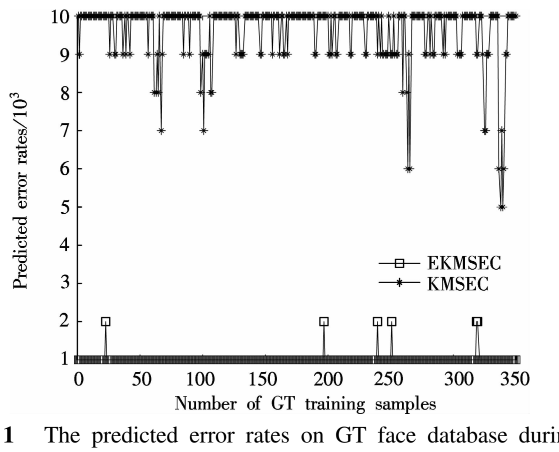

Fig. 1 The predicted error rates on GT face database during the training phase

3 Experiments

We employ FERET and GT face databases to test the presented EKMSE, KMSE and MNMSE[9]. Since the newly-immerging sparse-based representation method can be considered as a variant of KMSE, we choose CRC[4] for comparison. The Gaussian kernel function in the form of $K(x, y) = \exp(-\|x - y\|^2/2\delta^2)$ is adopted here, where δ^2 is set to be the variance of the first training partition. γ in Eq. (8) is 0.001. In EKMSE, the number of iterations is 10. The numbers in the first rows of Tab. 1 and Tab. 2 illustrate the training samples per class, and the remaining images are taken as test samples.

Tab. 1 Recognition accuracy of different methods on the FERET face database %

Methods	Number of original training samples per class		
	3	4	5
MSE	57.24	59.38	71.76
KMSE	60.65	67.62	73.13
MNMSE[9]	62.13	67.23	71.46
CRC[4]	49.12	58.68	67.67
EKMSE	62.31	69.96	75.80

Tab. 2 Recognition accuracy of different methods on the GT face database %

Methods	Number of original training samples per class				
	3	4	5	6	7
MSE	41.83	41.27	48.60	63.56	68.86
KMSE	52.06	56.71	60.27	63.06	65.35
MNMSE[9]	51.43	54.38	55.19	64.23	67.21
CRC[4]	50.50	53.45	56.40	64.44	68.25
EKMSE	52.47	57.32	61.04	64.00	66.44

The FERET face database[10] contains 1 400 gray images from 200 subjects. Every subject provides 7 frontal view face images with pose variations of ±15°, ±25°, and also the variations of the illumination and expression. Before the experiment, we adopt the down-sampling algorithm to resize each image into 40 × 40 pixels. Tab. 1 shows that our proposed method usually classifies more accurately than MSE, KMSE, and CRC, as well as the KMSE improvement, MNMSE.

We then use the Georgia Tech (GT) face database[11] to test our method. It contains images of 50 people taken in two or three sessions. All people in the database were represented by 15 color JPEG images with a cluttered background taken at the resolution of 640 × 480 pixels. The pictures show frontal or tilted faces with different facial expressions, lighting conditions and scales. Each image was manually labeled to determine the position of the face in the image. We use the face images with the background removed and each of these face images has the resolution of 40 × 30 pixels. They are all converted into gray images in advance. Tab. 2 shows that the EKMSE achieves higher recognition accuracy rates than KMSE, MNMSE and CRC.

4 Conclusion

In this paper, we develop an EKMSE algorithm to improve the classification performance of KMSE. We modify the objective function by adding an adaptive item to enlarge the distances of different classes. The experiments conducted on FERET and GT face databases demonstrate that the EKMSE can obtain more satisfactory classification performance than KMSE, MNMSE and CRC.

References

[1] Muller K, Mika S, Ratsch G, et al. An introduction to kernel-based learning algorithms [J]. *IEEE Transactions on Neural Networks*, 2001, **12**(2): 181 – 202. DOI: 10.1109/72.914517.

[2] Xu Jianhua, Zhang Xuegong, Li Yanda. Kernel MSE algorithm: a unified framework for KFD, LS-SVM and KRR [C]//*IEEE International Joint Conference on Neural Networks*. Washington, DC, USA, 2001: 1486 – 1491.

[3] Xu Yong, Zhang David, Yang Jian, et al. A two-phase test sample sparse representation method for use with face recognition [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, **21**(9): 1255 – 1262.

[4] Zhang Lei, Yang Meng, Feng Xiangchu. Sparse representation or collaborative representation: Which helps face recognition? [C]//*IEEE International Conference on Computer Vision*. Barcelona, Spain, 2011: 471 – 478.

[5] Qi Zhu. Reformative nonlinear feature extraction using kernel MSE [J]. *Neurocomputing*, 2010, **73**(16/17/18): 3334 – 3337. DOI: 10.1016/j.neucom.2010.04.007.

[6] Zhao Yongping, Sun Jianguo, Du Zhonghua, et al. Pruning least objective contribution in KMSE [J]. *Neurocomputing*, 2011, **74**(17): 3009 – 3018. DOI: 10.1016/j.neucom.2011.04.004.

[7] Zhao Yongping, Wang Kangkang, Liu Jie, et. al. Incremental kernel minimum squared error (KMSE) [J]. *Information Sciences*, 2014, **270**: 92 – 111. DOI: 10.1016/j.ins.2014.02.117.

[8] Xu Yong, Yang Jingyu, Jin Zhong, et al. A learning approach to derive sparse kernel minimum square error model [C]//*IEEE International Conference on Control and Automation*. Guangzhou, China, 2007: 1278 – 1283.

[9] Wang Jinhua. A novel solution scheme for the kernel MSE model[C]//*International Conference on Artificial Intelligence and Computational Intelligence*. Shanghai, China, 2009: 375 – 378.

[10] Counterdrug Technology Development Program. The FERET database[EB/OL]. (2004-06-16)[2016-01-30]. http://www.itl.nist.gov/iad/humanid/feret.

[11] Geogia Institute of Technology. Georgia Tech face database [EB/OL]. (2010-01-01) [2016-01-30]. http://www.anefian.com/research/face_reco.htm.

增强 KMSE 及人脸识别应用

赵英男[1,2,3] 何祥健[3] 陈北京[1,2] 赵晓平[1,2]

([1] 南京信息工程大学江苏网络监控工程中心,南京 210044)
([2] 南京信息工程大学计算机与软件学院,南京 210044)
([3] 悉尼科技大学计算机与通信学院,悉尼 2007)

摘要:为了提高核最小均方误差(KMSE)方法的识别能力,提出一种增强 KMSE 方法(EKMSE). 该方法重新定义 KMSE 目标函数,引入一个新的类别标签定义,并使该定义下的类别标签矩阵能够随核矩阵自适应调整. 与通常的目标函数相比,它能够使不同类别之间的距离增大,进而提高识别率. 同时该算法在参数搜索中采用了迭代技术,有效提高了算法的计算效率. 在 FERET 和 GT 人脸库上进行了充分的实验,结果表明 EKMSE 算法可行有效. 该算法不仅优于原 MSE, KMSE 以及 KMSE 改进算法,也优于目前脸识别中的基于稀疏算法的最新技术 CRC 算法.

关键词:最小均方误差;核最小均方误差;模式识别;人脸识别

中图分类号:TP391