Mean representation based classifier with its applications

J. Xu and J. Yang

Based on a fundamental concept that most similar properties of samples from a single-object class should be congregated on their class mean, an efficient and simple approach for pattern identification, called the mean representation based classifier (MRC), is presented. MRC is a linear model representing a testing sample as a linear combination of all class means and the class associating the biggest item of the linear combination coefficient is favoured. MRC is easy to employ with a least squares estimator. In addition, MRC need not tune any parameter and avoids mistaking the local optimum value as the global optimal one. MRC is evaluated on three standard databases. The experimental results show MRC is superior to other state-of-the-art nonparametric classifiers.

Introduction: The minimum distance classifier (MDC) has aroused broad interest in pattern recognition and computer vision areas [1]. MDC uses the class mean as the prototype of the class, and then the classification is done based on the distance from the testing sample to the prototype of each class. Thus this classifier is also called the nearest mean classifier. Let m_i $(i = 1, \ldots, c)$ be the mean vector of the training samples in the ith class. The square distance from $\mathbf{x}$ to the ith class is defined by $d_i(\mathbf{x}) = \|\mathbf{x} - \mathrm{m}_i\|^2$. If the distance between $\mathbf{x}$ and the sth class is minimal, i.e. $d_s(\mathbf{x}) = \min_i d_i(\mathbf{x})$, the decision of MDC is that $\mathbf{x}$ belongs to the sth class. The MDC method turns out to be very effective, especially for regular clustering data.

Observing the decision rule of MDC, a testing sample is assigned to the class whose mean is close to it. The excellent performance of MDC is based on the assumption that the data are well clustered and the samples with the same label lie on a common manifold, and simultaneously, have a similar orientation. This means the samples coming from the same class should have some similar properties. These similar properties are concentrated on the class mean. In other words, the most obvious common properties of the samples from the same class should manifest in the class mean.

Inspired by these advantages of MDC, many class-mean based techniques have been developed and have achieved satisfying performances. These excellent performances have also motivated us to develop a simple and efficient classifier, called the mean representation based classifier (MRC). MRC also takes the class mean as the prototype and uses the collection of overall class means as the prototype dictionary to find the optimal approximator of a testing sample and its corresponding optimal combination coefficient vector. According to the maximum similarity rule, the decision can be made in favour of the class with the biggest item of the optimal combination coefficient vector. The presented method can avoid the difficulties of parameter selection, which are the common difficulties in most parametric classifiers, such as the sparse representation classifier (SRC) [2] with the regularisation parameter selection, the local mean based nearest neighbour classifier (LM-NNC) [3] and the K nearest neighbours classifier (KNN) [4] both with the local neighbourhood parameter setting etc., despite their benefits in pattern classification. In this Letter, the classification problem is finally reformulated as the least squares estimator. Therefore it is easy to find the optimal linear combination coefficient by solving a least squares problem.

Mean representation based classifier: Suppose there are C known pattern classes. Let us define a matrix $X = [X_1, \cdots X_C] \in R^{d\times N}$, where X_i are the samples from the ith class, d is the dimension of sample in the input space and N is the number of total training samples. We define another matrix $\mathrm{M} = [\mathrm{m}_1, \mathrm{m}_2, \ldots, \mathrm{m}_c] \in R^{d\times N}$, called the class mean matrix, where m_i is the mean vector of the ith class.

Given an unlabelled test sample $\mathrm{y} \in R^d$, our problem is to assign the label to it. Now, we represent y as the class mean vectors, i.e. $\mathrm{y} = \mathrm{Mw}$, where $\mathrm{w} = [w_1, \cdots, w_C]$ is the coefficient vector. If y belongs to the sth class, the corresponding item of coefficient vector w should be the biggest one among the C items of coefficient vector w.

Given that $d \geq C$, the system of linear equations $\mathrm{y} = \mathrm{Mw}$ is well conditioned and thus very stable. The coefficient vector w can be obtained by solving the least square optimisation problem [5] and the solution is

$$\mathrm{w} = (\mathrm{M}^T\mathrm{M})^{-1}\mathrm{M}^T\mathrm{y}$$

Now we find the biggest one from C items of the obtained optimal coefficient vector w; if

$$w_s = \max_i (w_i)$$

the testing sample y will be assigned to the sth class.

Advantages of MRC: First, without parameter tuning, MRC can save time-costing and, more importantly, avoids the embarrassment that happens in many parametric classifiers of mistaking the local optimum value as the global optimal one. Secondly, the key to MRC is to solve the least squares optimisation problem, which is easy to implement with the computational complexity of $o(dC^2)$. Finally, the least squares optimisation system of MRC is stable in dealing with the SSS problems, since $d \geq N \geq C$ such that $\mathrm{M}^T\,\mathrm{M}$ is nonsingular.

Experiments and analyses: Database descriptions and comparative classifiers: The performance of the MRC approach is evaluated on the PolyU palmprint database (http://www4.comp.polyu.edu.hk/~biometrics/), whose data are all preprocessed using histogram equalisation, the Wine database from UCI (http://archive.ics.uci.edu/ml) and the Yale face database (http://cvc.yale.edu). Detailed descriptions of the three databases are listed in Table 1. We compare MRC with the three nonparametric classifiers: the nearest neighbour classifier (NNC) [6], the minimum distance classifier (MDC) [1] and the linear regression classifier (LRC) [7]. In addition, note that in our method each image vector is normalised so that the maximum pixel value is 1.

Table 1: Details of three databases

Database	Dimensions	Class	Size of sample per class
PolyU	64 × 64	100	6
Wine	13	3	48
Yale	100 × 80	15	11

Experimental settings and results: In the PolyU palmprint database, the three images captured in the first session are used to find the PCA subspace and the last three images captured in the second session are used for testing. The experiment with reverse order is employed again. Both experiments are performed on the 90-dimensional PCA subspaces. In the Wine database from UCI, we randomly select 10 and 15 samples per class for training and repeat experiments 20 times on the original 13-dimensional features, and finally report the average maximum recognition rates of each methods. In the Yale face database, the first three face images are selected to find the 40-dimensional PCA transformed space and the remaining eight are used to test. All experimental results are listed in Table 2.

Table 2: Recognition rates (%) and corresponding dimension (·) or standard deviations (±·) of different algorithms on three datasets

		NNC	MDC	LRC	MRC
PolyU	First 3	86.0(90)	84.3(65)	86.3(80)	98.7(85)
	Last 3	78.3(75)	77.7(90)	81.7(85)	97.3(70)
Wine	10	64.6(±4.2)	65.7(±3.2)	76.5(±7.8)	79.8(±3.5)
	15	64.6(±4.8)	66.6(±3.4)	33.3(±0.0)	79.4(±4.7)
Yale	First 3	90.0(18)	90.0(22)	94.2(32)	95.0(38)

Experimental analyses: The experimental results listed in Table 2 clearly reflect the potency of MRC in dealing with the SSS problem. Especially, on the PolyU palmprint databases, we find the difference between MRC and other comparative methods becomes particularly significant. By comparing the recognition results obtained on the Wine database, we find MRC performs best among four comparative methods irrespective of the variation in the size of the training sample in each class. In contrast, LRC is eclipsed. When the size of the training sample per class increases from 10 to 15, the performance of LRC degrades a lot with a lowest recognition rate of 33.3%. The main reason is that LRC is specially designed for face recognition, which requires the size of the sample set per class to be smaller than the dimension of the sample. When 15 (>13) samples per class is used for training, the recognition system of LRC is out of work. Fortunately, MRC does not suffer this difficulty even using the same least squares estimator as LRC. The main reason is that the feature dimension is higher than the number of classes in the Wine database, i.e. $13 > 3$. Thus the least squares estimator in MRC still works. This indicates that only if the number of classes is less than the dimensionality of the sample can MRC do well, even in the large sample size case. Actually, in the SSS problem, the dimension of the sample must be larger than the number of classes. From this point of view, MRC is more robust than LRC, since besides the SSS problem, MRC can deal with more cases only if the dimension of the sample is larger than the number of classes.

Conclusions: In this Letter, a novel and simple classification method, called the mean representation based classifier (MRC), is proposed which formulates the pattern identification task as a problem of least squares estimator. MRC is a nonparametric classification, thus it is easy to implement without tuning any parameter. In addition, MRC can work well in many cases only if the dimension of the class is larger than the number of classes. Experimental results on three publicly available databases demonstrate the effectiveness of our proposed method in pattern identification.

31 July 2011
doi: 10.1049/el.2011.2420

J. Xu and J. Yang (*Department of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, People's Republic of China*)

E-mail: njxujie@yahoo.cn

J. Yang: Also with Computation and Neural Systems, California Institute of Technology, Pasadena, CA 91125, USA

References

1 Gonzalez, R.C., and Woods, R.E.: 'Digital image processing' (Addison Wesley, 1997)
2 Wright, J., Yang, A., Ganesh, A., Sastry, S., and Ma, Y.: 'Robust face recognition via sparse representation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009, **31**, (2), pp. 210–227
3 Yang, J., Zhang, L., Yang, J.Y., and Zhang, D.: 'From classifiers to discriminators: a nearest neighbor rule induced discriminant analysis', *Pattern Recognit.*, 2011, **44**, (7), pp. 1387–1402
4 Fukunaga, K.: 'Introduction to statistical pattern recognition' (Academic Press, 1990, 2nd edn.)
5 Hastie, T., Tibshirani, R., and Friedman, J.: 'The elements of statistical learning; data mining, inference and prediction' (Springer, 2001)
6 Cover, T.M., and Hart, P.E.: 'Nearest neighbor pattern classification', *IEEE Trans. Inf. Theory*, 1967, **13**, (1), pp. 21–27
7 Naseem, I., Togneri, R., and Bennamoun, M.: 'Linear regression for face recognition', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010, **32**, (11), pp. 2106–2112