LETTER

Multi-View Synthesis and Analysis Dictionaries Learning for Classification

Fei WU[†a)], ***Member***, **Xiwei DONG**[†], **Lu HAN**[†], **Xiao-Yuan JING**[†], ***and*** **Yi-mu JI**[††b)], ***Nonmembers***

SUMMARY Recently, multi-view dictionary learning technique has attracted lots of research interest. Although several multi-view dictionary learning methods have been addressed, they can be further improved. Most of existing multi-view dictionary learning methods adopt the l_0 or l_1-norm sparsity constraint on the representation coefficients, which makes the training and testing phases time-consuming. In this paper, we propose a novel multi-view dictionary learning approach named multi-view synthesis and analysis dictionaries learning (MSADL), which jointly learns multiple discriminant dictionary pairs with each corresponding to one view and containing a structured synthesis dictionary and a structured analysis dictionary. MSADL utilizes synthesis dictionaries to achieve class-specific reconstruction and uses analysis dictionaries to generate discriminative code coefficients by linear projection. Furthermore, we design an uncorrelation term for multi-view dictionary learning, such that the redundancy among synthesis dictionaries learned from different views can be reduced. Two widely used datasets are employed as test data. Experimental results demonstrate the efficiency and effectiveness of the proposed approach.

key words: *multi-view dictionary learning, synthesis dictionary, analysis dictionary, uncorrelation term*

1. Introduction

Dictionary learning is an important feature learning technique with state-of-the-art classification performance. The Fisher discrimination dictionary learning method [1] learns a structured dictionary whose atoms correspond to the class labels. Gu et al. [2] presented a dictionary pair learning (DPL) method, which learns a synthesis dictionary and an analysis dictionary jointly to achieve the goal of signal representation and discrimination. Most of dictionary learning methods are developed to solve single view based learning problems.

In many real-world applications, the same object can be observed at different viewpoints or described with different descriptors, thus generating multi-view data. Since more useful information exists in multiple views than in a single one, multi-view learning has received significant attention. The multi-view canonical correlation analysis (MCCA) method [3] aims to exploit the correlation features among multiple views. The multi-view discriminant analysis (MvDA) method [4] minimizes the within-class variations and maximizes the between-class variations of samples in the learning common space from both intra-view and inter-view.

Recently, some multi-view dictionary learning methods have been presented and achieved impressive classification performance. The supervised coupled dictionary learning with group structures for multi-modal retrieval (SliM^2) method [5] utilizes the class label information of samples to jointly learn discriminative multi-modal dictionaries and mapping functions between different modalities. The multimodal sparse representation-based classification (MSRC) method [6] selects the topmost discriminative samples for each individual modality, guaranteeing the large diversity among different modalities, for classifying lung needle biopsy images. Jing et al. [7] developed an uncorrelated multi-view discrimination dictionary learning (UMD^{2}L) method to jointly learn multiple uncorrelated discriminative dictionaries from multiple views. Wang et al. [8] presented a multi-view analysis dictionary learning (MvADL) method, which learns multi-view analysis dictionaries and utilizes a marginalized classification term to integrate the label information into the dictionary learning model.

Most of existing multi-view dictionary learning methods (except for MvADL [8]) mainly focus on learning synthesis dictionaries [10] to represent the input signal. And they usually utilize l_0-norm or l_1-norm to regularize the coding coefficient. However, the use of l_0 or l_1-norm sparsity regularization will lead to large computational cost. For MvADL, it only focuses on analysis dictionary learning without taking synthesis dictionary learning into consideration. As stated in [9], analysis dictionary directly transforms a signal to a feature space by multiplying the signal, while synthesis dictionary represents an input signal by using a linear combination of dictionary atoms, which provides a complementary view of data representation. In addition, MvADL also adopts the l_0-norm sparsity regularizer on coding coefficient, which needs relatively large computational cost. How to effectively improve the efficiency of multi-view dictionary learning technique but preserve its effectiveness is an important research topic.

Inspired by single view based dictionary learning work, i.e., DPL [2], which learns discriminative synthesis and analysis dictionary pair and uses the analysis coding scheme, we propose a multi-view synthesis and analysis dictionaries learning (MSADL) approach. We summarize the

Manuscript received May 23, 2018.
Manuscript revised September 28, 2018.
Manuscript publicized November 27, 2018.

[†] The authors are with the College of Automation, Nanjing University of Posts and Telecommunications (NJUPT), Nanjing 210003, China.

[††] The author is with the College of Computer, NJUPT, Nanjing 210003, China.

a) E-mail: wufei_8888@126.com
b) E-mail: jiym@njupt.edu.cn

DOI: 10.1587/transinf.2018EDL8107

contributions of our work as following three points:

(1) We provide a new multi-view dictionary learning idea that is jointly learning a pair of dictionaries (a structured synthesis dictionary and a structured analysis dictionary) for each view. We propose to use synthesis dictionary to achieve class-specific reconstruction and utilize analysis dictionary to approximate discriminative representation coefficients by linear projection. Since we employ an analytical coding scheme, the efficiency in both training and testing phases can be effectively improved. The structured dictionary learning manner can also bring about dictionaries with favorable discriminative power.

(2) We design an uncorrelation term for multi-view dictionary learning, so as to reduce the redundancy among synthesis dictionaries of different views.

(3) We verify the proposed MSADL approach on the Multi-PIE [11] and the MNIST [12] datasets. Experimental results demonstrate its efficiency and effectiveness as compared with several related methods.

2. Proposed Approach

2.1 The Model of MSADL

Assume that there exists M views of training data, denoted by $A_k = [A_k^1, \cdots, A_k^C] \in \mathbb{R}^{l\times Cn}$ $(k = 1, \cdots, M)$, where $A_k^i \in \mathbb{R}^{l\times n}$ is the training sample subset from the i^{th} class. Here, l and n separately represent the dimensionality and number of samples in A_k^i, and C denotes the number of classes. We aim to learn a structured synthesis dictionary $D_k = [D_k^1, \cdots, D_k^C] \in \mathbb{R}^{l\times Cq}$ and a structured analysis dictionary $P_k = [P_k^1; \cdots; P_k^C] \in \mathbb{R}^{Cq\times l}$ for each view. For the k^{th} view, $D_k^i \in \mathbb{R}^{l\times q}$ and $P_k^i \in \mathbb{R}^{q\times l}$ separately denote the class-specified synthesis and analysis sub-dictionaries associated with class i. For simplicity, q is set to be equal to n in this paper.

First of all, since the synthesis sub-dictionary D_k^i is associated with the i^{th} class, we require that it should well reconstruct A_k^i with the corresponding projective coding coefficient matrix $P_k^i A_k^i$. Thus, we should minimize the value of $\sum_{k=1}^{M}\sum_{i=1}^{C}\left\|A_k^i - D_k^i P_k^i A_k^i\right\|_F^2$. Second, we want that the analysis sub-dictionary P_k^i can project the samples from class $j\,(j \neq i)$ to a nearly null space, i.e., $P_k^i A_j \approx 0, \ \forall j \neq i$. We thus require the minimization of $\sum_{k=1}^{M}\sum_{i=1}^{C}\left\|P_k^i \bar{A}_k^i\right\|_F^2$, where $\bar{A}_k^i \in \mathbb{R}^{l\times(C-1)n}$ denotes the complementary data matrix of A_k^i in A_k. Third, to reduce the redundancy among synthesis dictionaries among different views, we provide an uncorrelation term that makes each synthesis sub-dictionary D_k^i be independent from $D_l\,(l \neq k)$. Correspondingly, the value of $\sum_{k=1}^{M}\sum_{i=1}^{C}\sum_{l=1,l\neq k}^{M}\left\|D_l^T D_k^i\right\|_F^2$ should be as small as possible.

According to above considerations, we formulate the objective function of MSADL as follows:

$$\begin{aligned}\arg\min_{\substack{P_1,\cdots,P_M\\D_1,\cdots,D_M}} &\sum_{k=1}^{M}\sum_{i=1}^{C}\Bigg(\left\|A_k^i - D_k^i P_k^i A_k^i\right\|_F^2 + \alpha\left\|P_k^i \bar{A}_k^i\right\|_F^2 \\ &\qquad + \beta\sum_{l=1,l\neq k}^{M}\left\|D_l^T D_k^i\right\|_F^2\Bigg) \qquad (1)\\ s.t.\ &\left\|d_k^m\right\|_2^2 \leq 1\end{aligned}$$

where α and β are balance factors, $d_k^m \in \mathbb{R}^l$ denotes the m^{th} atom (vector) in D_k. We constrain the energy of each dictionary-atom to avoid the trivial solution of $P_k^i = 0$ and make MSADL more stable. Mathematically, $\left\|D_l^T D_k^i\right\|_F^2 \leq \|D_l\|_F^2 \left\|D_k^i\right\|_F^2$. To make the above problem more tractable, we relax (1) into

$$\begin{aligned}\arg\min_{\substack{P_1,\cdots,P_M\\D_1,\cdots,D_M}} &\sum_{k=1}^{M}\sum_{i=1}^{C}\Bigg(\left\|A_k^i - D_k^i P_k^i A_k^i\right\|_F^2 \\ &\qquad + \alpha\left\|P_k^i \bar{A}_k^i\right\|_F^2 + \beta h_k^i\left\|D_k^i\right\|_F^2\Bigg) \qquad (2)\\ s.t.\ &\left\|d_k^m\right\|_2^2 \leq 1\end{aligned}$$

where $h_k^i = \sum_{l=1,l\neq k}^{M}\|D_l\|_F^2$.

2.2 Optimization and Classification

The objective function in Formula (2) is generally non-convex. We introduce variables $X_k = \{X_k^1, \cdots, X_k^C\}$ $(k = 1, \cdots, M)$, where $X_k^i \in \mathbb{R}^{q\times n}$ is a sub-matrix of X_k, and relax (2) to the following problem:

$$\begin{aligned}\arg\min_{\substack{P_1,\cdots,P_M\\X_1,\cdots,X_M\\D_1,\cdots,D_M}} &\sum_{k=1}^{M}\sum_{i=1}^{C}\Bigg(\left\|A_k^i - D_k^i X_k^i\right\|_F^2 + \gamma\left\|P_k^i A_k^i - X_k^i\right\|_F^2 \\ &\qquad + \alpha\left\|P_k^i \bar{A}_k^i\right\|_F^2 + \beta h_k^i\left\|D_k^i\right\|_F^2\Bigg) \qquad (3)\\ s.t.\ &\left\|d_k^m\right\|_2^2 \leq 1\end{aligned}$$

where γ is a balance factor. Then, we employ a two-level optimization strategy to update the variables, that is: 1) updating variables of the k^{th} view by fixing variables of the other views; 2) for the k^{th} view, updating X_k and $\{D_k, P_k\}$ alternatively.

For the k^{th} view, we firstly fix D_k and P_k to compute X_k. (3) can be reduced to

$$X_k^* = \arg\min_{X_k}\sum_{i=1}^{C}\left(\left\|A_k^i - D_k^i X_k^i\right\|_F^2 + \gamma\left\|P_k^i A_k^i - X_k^i\right\|_F^2\right) \qquad (4)$$

The solution of (4) can be analytically derived as

$$X_k^i = \left({D_k^i}^T D_k^i + \gamma I\right)^{-1}\left(\gamma P_k^i A_k^i + {D_k^i}^T A_k^i\right) \qquad (5)$$

where $I \in \mathbb{R}^{q\times q}$ is an identity matrix.

We then fix D_k and X_k to update P_k. We should solve the following problem:

$$P_k^* = \arg\min_{P_k} \sum_{i=1}^{C} \left(\gamma \left\| P_k^i A_k^i - X_k^i \right\|_F^2 + \alpha \left\| P_k^i \bar{A}_k^i \right\|_F^2 \right) \tag{6}$$

P_k can be easily obtained

$$P_k^{i*} = \gamma X_k^i {A_k^i}^T \left(\gamma A_k^i {A_k^i}^T + \alpha \bar{A}_k^i {\bar{A}_k^i}^T + \delta I \right)^{-1} \tag{7}$$

where δ is a small constant and can be set as $\delta = 10e^{-4}$.

When updating D_k, X_k and P_k are fixed. We should solve the following problem:

$$\begin{gathered} D_k^* = \arg\min_{D_k} \sum_{i=1}^{C} \left(\left\| A_k^i - D_k^i X_k^i \right\|_F^2 + \beta h_k^i \left\| D_k^i \right\|_F^2 \right) \\ s.t. \ \left\| d_k^m \right\|_2^2 \le 1 \end{gathered} \tag{8}$$

We introduce a variable matrix $Q_k \in \mathbb{R}^{l \times Cq}$. Then

$$\begin{gathered} D_k^* = \arg\min_{D_k, Q_k} \sum_{i=1}^{C} \left(\left\| A_k^i - D_k^i X_k^i \right\|_F^2 + \beta h_k^i \left\| D_k^i \right\|_F^2 \right) \\ s.t. \ D_k = Q_k, \ \left\| q_k^m \right\|_2^2 \le 1 \end{gathered} \tag{9}$$

where $q_k^m \in \mathbb{R}^l$ represents the m^{th} column in Q_k. The optimal solution of Formula (9) can be achieved by using the alternating direction method of multipliers (ADMM) algorithm [13]:

$$\begin{cases} D_k^{(t+1)} = \arg\min_{D_k} \sum_{i=1}^{C} \left(\left\| A_k^i - D_k^i X_k^i \right\|_F^2 + \beta h_k^i \left\| D_k^i \right\|_F^2 \right. \\ \qquad\qquad \left. + \eta \left\| D_k^i - Q_k^{i(t)} + T_k^{i(t)} \right\|_F^2 \right) \\ Q^{(t+1)} = \arg\min_{Q_k} \sum_{i=1}^{C} \eta \left\| D_k^{i(t+1)} - Q_k^i + T_k^{i(t)} \right\|_F^2, \\ \qquad\qquad s.t. \ \left\| q_k^i \right\|_2^2 \le 1 \\ T^{(t+1)} = T^{(t)} + D_k^{i(t+1)} - Q_k^{(t+1)} \end{cases} \tag{10}$$

When synthesis dictionaries $\{D_1, \cdots, D_M\}$ and analysis dictionaries $\{P_1, \cdots, P_M\}$ are obtained, a test sample $y = \{y_1, y_2, \cdots, y_M\}$ can be classified via coding it over these dictionaries. Here, $y_k \in \mathbb{R}^l$ denotes a vector of the k^{th} view of y. The synthesis sub-dictionary D_k^i is trained to well reconstruct samples from the i^{th} class rather than the j^{th} $(j \neq i)$ class. In addition, the analysis sub-dictionary P_k^i is trained to generate significant coefficients for samples from the i^{th} class and small coefficients for samples from classes other than i. Therefore, if y_k $(k = 1, \cdots, M)$ is from the i^{th} class, its projective coding vector by P_k^i, i.e., $P_k^i y_k$, tends to be significant, while its projective coding vectors by P_k^j $(j \neq i)$ tend to be small. And the reconstruction error $\left\| y_k - D_k^i P_k^i y_k \right\|_2^2$ will be smaller than $\left\| y_k - D_k^j P_k^j y_k \right\|_2^2$. Then, we can employ the class-specific reconstruction error to classify test samples. We define the metric for final classification as:

$$e_i = \sum_{k=1}^{M} \left\| y_k - D_k^i P_k^i y_k \right\|_2^2 \tag{11}$$

We do classification via $identity\,(y) = \arg\min_i \{e_i\}$.

3. Experiments

3.1 Compared Methods and Experimental Settings

We compare the proposed approach with the representative multi-view subspace learning methods MCCA [3] and MvDA [4], and four state-of-the-art multi-view dictionary learning methods, i.e., SliM^2 [5], MSRC [6], UMD^{2}L [7] and MvADL [8] on the widely used Multi-PIE [11] and MNIST [12] datasets.

In all experiments, the tuning parameters in MSADL (α, β, γ and η) are evaluated by 5-fold cross validation on training samples. Concretely, these parameters are set as $\alpha = 0.003, \beta = 0.0005, \gamma = 0.05$ and $\eta = 1$ for the Multi-PIE dataset; and $\alpha = 0.003, \beta = 0.01, \gamma = 0.05$ and $\eta = 1$ on the MNIST dataset.

3.2 Introduction of Datasets

In experiment, we employ a subset of the Multi-PIE dataset [11] for experiment, which contains face image samples from 68 peoples. In this dataset, there exist 24 samples for each people with 5 different poses (C05, C07, C09, C27, C29). The image size is 64 × 64 pixels. Like [14], principal component analysis (PCA) transformation [15] is used to reduce the dimension of samples to 100. We randomly select 8 samples (each sample with 5 different poses) per class for training and use the remaining samples for testing. We repeat random selection 20 times to report the average results.

The MNIST dataset [12] used in the experiment contains 1000 handwritten digit images (100 images for each digit). The image size is 28 × 28 pixels. We extract the Gabor transformation, KL transformation and LBP features to build three feature sets for experiment. We also employ PCA to reduce the dimension of feature samples to 100. We randomly select 40 samples per class for training, use the remaining samples for testing, and run compared methods 20 times.

3.3 Experimental Results

We list the average classification accuracies of all compared methods on the Multi-PIE and MNIST datasets in Table 1. We can see that the proposed approach achieves the best classification results as compared with competing methods on the Multi-PIE dataset. Specifically, MSADL improves the average classification accuracies at least by 0.17% (= 96.02% − 95.85%) on the Multi-PIE dataset. With respect to the MNIST dataset, MSADL outperforms multi-view subspace learning methods MCCA and MvDA, and multi-view dictionary learning methods including SliM^2, MSRC and MvADL, and obtains the comparable classification results as compared with UMD^{2}L.

Table 1 Average classification accuracies (CA, %) and running time (seconds) of compared methods.

Method	Multi-PIE		MNIST	
	CA	Time	CA	Time
MCCA	94.06 ± 1.09	1.48	86.76 ± 1.23	0.31
MvDA	94.28 ± 0.92	7.47	87.42 ± 1.17	1.26
SliM^2	94.47 ± 0.12	287.24	88.51 ± 0.98	59.22
MSRC	93.02 ± 1.31	261.31	89.52 ± 1.54	55.79
UMD^{2}L	95.85 ± 1.02	3476.16	**90.44** ± 1.26	912.87
MvADL	95.71 ± 0.93	16.96	90.20 ± 1.22	3.75
MSADL	**96.02** ± 0.64	**6.79**	90.43 ± 0.81	**0.91**

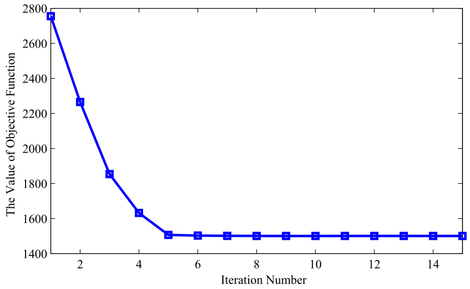

Fig. 1 Convergence effect of MSADL on Multi-PIE.

Table 1 also shows running time of all compared methods. Our hardware configuration comprises 32-bit computers with 2.09-GHz dual core processer and 4GB RAM. It is obvious that our proposed approach costs significantly less running time than that of existing multi-view dictionary learning methods, which demonstrates the efficiency of MSADL. From the table, we can conclude that MSADL effectively improves the efficiency of multi-view dictionary learning technique without sacrificing the classification effect.

Figure 1 shows the convergence effect of MSADL on the Multi-PIE dataset. We can see that the value of objective function of MSADL decreases sharply as iterations increase, which indicates that our approach can converge rapidly. On the MNIST dataset, MSADL can also converge with a few iterations.

4. Conclusions

In this paper, to improve the computational efficiency of multi-view dictionary learning technique, we propose a novel approach named MSADL. It learns a set of structured synthesis and analysis dictionaries from multiple views. With these synthesis and analysis dictionaries, MSADL performs representation and discrimination simultaneously. In addition, MSADL provides an uncorrelation term to reduce the redundancy among synthesis dictionaries. Experiments on two public datasets demonstrate that MSADL effectively improves the efficiency of multi-view dictionary learning technique without sacrificing the classification accuracy.

Acknowledgments

The work in this paper was supported by the National Natural Science Foundation of China (No. 61702280), National Postdoctoral Program for Innovative Talents (No. BX20180146), Natural Science Foundation of Jiangsu Province (No. BK20170900), Natural Science Fund for Colleges and Universities in Jiangsu Province (No. 17KJB520025), and Scientific Research Staring Foundation for Introduced Talents in NJUPT (NUPTSF, No. NY217009).

References

[1] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse representation based Fisher discrimination dictionary learning for image classification," International Journal of Computer Vision, vol.109, no.3, pp.209–232, 2014.

[2] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Projective dictionary pair learning for pattern classification," Advances in Neural Information Processing Systems, pp.793–801, 2014.

[3] Y.-O. Li, T. Adali, W. Wang, and V.D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," IEEE Trans. Signal Process., vol.57, no.10, pp.3918–3929, 2009.

[4] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol.38, no.1, pp.188–194, 2016.

[5] Y. Zhuang, Y. Wang, F. Wu, Y. Zhang, and W. Lu, "Supervised coupled dictionary learning with group structures for multi-modal retrieval," AAAI Conference on Artificial Intelligence, pp.1070–1076, 2013.

[6] Y. Shi, Y. Gao, Y. Yang, Y. Zhang, and D. Wang, "Multimodal sparse representation-based classification for lung needle biopsy images," IEEE Trans. Biomed. Eng., vol.60, no.10, pp.2675–2685, 2013.

[7] X.Y. Jing, R. Hu, F. Wu, X. Chen, Q. Liu, and Y.F. Yao, "Uncorrelated multi-view discrimination dictionary learning for recognition," AAAI Conference on Artificial Intelligence, pp.2787–2795, 2014.

[8] Q. Wang, Y. Guo, J. Wang, X. Luo, and X. Kong, "Multi-view analysis dictionary learning for image classification," IEEE Access, vol.6, pp.20174–20183, 2018.

[9] M. Yang, H. Chang, and W. Luo, "Discriminative analysis-synthesis dictionary learning for image classification," Neurocomputing, vol.219, pp.404–411, 2017.

[10] P. Sprechmann, R. Litman, T.B. Yakar, A. Bronstein, and G. Sapiro, "Efficient supervised sparse analysis and synthesis operators," Advances in Neural Information Processing Systems, pp.908–916, 2013.

[11] D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal Laplacianfaces for face recognition," IEEE Trans. Image Process., vol.15, no.11, pp.3608–3614, 2006.

[12] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE, vol.86, no.11, pp.2278–2324, 1998.

[13] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," Foundations and Trends in Machine Learning, vol.3, no.1, pp.1–122, 2011.

[14] F. Wu, X.-Y. Jing, X. You, D. Yue, R. Hu, and J.-Y. Yang, "Multi-view low-rank dictionary learning for image classification," Pattern Recognit., vol.50, pp.143–154, 2016.

[15] M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neuroscience, vol.3, no.1, pp.71–86, 1991.