Sparse Representation for Classification with Structure Preserving Dimension Reduction

Jin Xu JXU4@STEVENS.EDU
Guang Yang GYANG1@STEVENS.EDU
Hong Man HONG.MAN@STEVENS.EDU
Electrical and Computer Engineering Department, Stevens institute of Technology, Hoboken, NJ 07030 USA

Abstract

Sparse representation based classifier (SRC), which is via reconstructed error, has been a popular method for recognition tasks. However, the computation cost for sparse coding is heavy with high dimensional datasets. In this paper, structure preserving with various dimension reduction methods are studied in the context of SRC to improve computational efficiency as well as classification accuracy. The inner relations among data, such as principle components, Pearson's correlations, and Laplacian scores, are applied to preserve the structure in low dimension. Comparisons of classification performance are made among SRC with structure preserving dimension reduction (SRC-SPDR) and classical classifiers such as SVM and k-nearest neighbors (KNN). Experimental results on UCI datasets and face databases show that SRC-SPDR can achieve competitive accuracy at relatively low computation cost.

1. Introduction

Sparse representation (or sparse coding) has been studied extensively in recent years. The key idea is to use the least number of basis vectors (or atoms) in a dictionary for describing a signal. It is well known that sparse coding methods are generally heavy in computation. Normally, there are three approaches to speed up the sparse coding process. (1) Structure preserving dimensionality reduction: The intention is to remove data redundancy at the data preparing step. Various dimension reduction methods can be applied on the data to obtain meaningful structure in low dimension, which is the focus of this paper. (2) Dictionary learning (Mairal et al., 2009): Various technologies such as regularization and clustering can be applied on data to train the dictionary. (3) Efficient algorithm: The approach is to apply different optimization methods to speed up sparse coding process.

Submission in *the 28th International Conference on Machine Learning Workshop*, Bellevue, WA, USA, 2011.

In this paper, we investigate a combined sparse representation based classifier (SRC) and structure preserving dimension reduction (SPDR) framework, and attempt to show that SRC can be effectively integrated with SPDR to achieve competitive performance. Dimension reduction can effectively extract useful structure and reduce the computation cost, which contributes to better classification and recognition performance (Lacoste-Julien et al., 2008). In (Wright et al., 2009), Randomfaces were used for structure preserving dimension reduction and in face recognition. In this work, three dimension reduction methods, i.e. PCA, Laplacian score feature selection (LAP), and Pearson's correlation coefficient (COR) feature selection, are studied in the SRC-SPDR framework, and extensive experiments in comparison with other classic classifiers (SVM and KNN) are conducted.

2. SRC Based on Structure Preserving Dimension Reduction

In sparse representation, given a dictionary containing a set of training data vectors (or atoms) $\mathbf{A} = [a_1^1, \cdots, a_1^{n_1}, \cdots, a_c^1, \cdots, a_c^{n_c}]$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$, c is the category label for each atom, n_i is the number of atoms associated with the category i. The goal is to represent a new test data vector y in the form

$$y = Ax \in \mathbb{R}^m \tag{1}$$

where $x = [0, \cdots, 0, \alpha_{i,1}, \alpha_{i,2}, \cdots, \alpha_{i,n_i}, 0, \cdots, 0]^T \in \mathbb{R}^n$ is the sparse vector. To solve this problem, ℓ_1-regularized least squares problem (Tibshirani, 1996), (Kim et al., 2007) is defined by

$$\widehat{x} = arg\min\{\|y - Ax\|_2^2 + \lambda\|x\|_1\} \quad (2)$$

An SRC may exploit the representation residual to identify the target category (Wright et al., 2009). For each category i, a characteristic function is defined as $\delta_i : \mathbb{R}^n \to \mathbb{R}^n$, which just selects the coefficients associated to i-th category. Then the classification is based on:

$$label(y) = arg\min r_i(y), \quad r_i(y) = \|y - A\delta_i(x)\|^2 \quad (3)$$

Because this optimization problem is computational intensive, it is highly desirable to reduce the dimensions of $\mathbf{A}$ and y. A dimension reduction matrix $P^{d\times m}$ is applied to the input y and the atoms of $\mathbf{A}$. Algorithm 1 shows detail procedure of the sparse representation based classification method with dimension reduction (SRC-SPDR).

Algorithm 1 SRC with SPDR

1: **Input:** a set of training data $A \in \mathbb{R}^{m\times n}$ with c categories, a test data $y \in \mathbb{R}^m$, a dimension reduction matrix $P \in \mathbb{R}^{d\times m}$
2: Compute $\tilde{A} = PA$ and $\tilde{y} = Py$
3: Solve ℓ_1-regularized least squares problem: $\widehat{x} = arg\min\{\|\tilde{y} - \tilde{A}x\|_2^2 + \lambda\|x\|_1\}$
4: Compute the residuals: $r_i(y) = \|\tilde{y} - \tilde{A}\delta_i(x)\|^2 \quad for \quad i = 1, \cdots, c$
5: **Output:** $label(y) = arg\min r_i(y)$

PCA, LAP and COR are dimension reduction techniques used in this work to discover the valid low dimensional structure. The PCA is a linear data transform, which is usually used to preserve the global structure by projecting the data to the directions of maximal variances. Laplacian scores (LAP) (He et al., 2005) reflect the geometric structure in nearest neighbor graph. This local structure is constructed by the weights between two nodes, connected or not. Hence, this structure preserves the discriminate features in the feature space. Pearson's correlation coefficient (COR) (Guyon & Elisseeff, 2003) is a supervised feature selection method. COR method constructs sub-feature space with features most concentrated on class center while filters out features less correlated, which is a criteria of dependence structure between the features and the classification labels.

3. Experimental Results

The datasets are from UCI datasets (Frank & Asuncion, 2010) and the extended Yale face database B. we apply three dimension reduction methods to transfer data to lower dimensional space. Then use sparse representation based classification (SRC), support vector machine (SVM), and k-nearest neighbor (KNN) methods to obtain the classification results. Each dataset is divided randomly into training dataset and testing dataset with the rate 1:1. Each experiment is repeated for five times, and the results are averaged. In our SRC, the dictionary contains the entire training dataset, which is similar to several recent successful works (Wright et al., 2009) (Yang & Zhang, 2010). The l_1-ls sparse coding software from Stanford (Kim et al., 2007) is used in the experiments. The SVM and KNN classification tools are from the Java toolbox (Witten & Frank, 2002), and parameters of the tools are kept in default.

Table 1. UCI Experiment Data Sets

Name	Feature number	Total size	Class number
Wine	13	178	3
Glass	10	214	7
Libras Movement	90	360	15
Wine Quality	11	4898	6
Anneal(△)	11	798	2

△ the missing feature has been removed.

3.1. Experiments on UCI data sets

Five UCI datasets (Frank & Asuncion, 2010) are selected in the experiments. We are more interested in multi-category classification problem, which is typically more challenging. Most selected datasets are multi-category datasets (except "Anneal"). Details of these datasets are described in Table 1. In the experiments, the dimensions of the feature sets are reduced to different sizes. The dimensional size are adjusted from small to large to evaluate the different effects on the classification.

In Figure 1, SRC shows obvious advantage on "Libras Movement" dataset. SRC can achieve around 20% higher accuracy compared with SVM and KNN in three reduced features. As the original datasets has 90 features. SRC can reach around 75% accuracy using 20 features, which is even higher than other classifiers using 80 features. We also notice that SRC with PCA can maintain the accuracy while the accuracies of SVM with PCA and KNN with PCA decrease at higher dimensions. This means that SRC may have the ability to deal with noise data.

Table 2 shows the comprehensive results based on UCI data sets. The classification accuracies of SVM, KNN and SRC are presented in different dimensions (20%,

Table 2. Comparisons of classification accuracy(%) based on UCI datasets

Data set	PCA 20%	PCA 40%	PCA 60%	PCA 80%	LAP 20%	LAP 40%	LAP 60%	LAP 80%	COR 20%	COR 40%	COR 60%	COR 80%
Wine-SVM	62.92	**95.51**	96.63	**97.75**	**78.65**	82.02	88.76	93.26	42.70	**92.13**	**89.89**	91.01
Wine-KNN	**68.54**	92.13	87.64	91.01	77.53	**83.15**	93.26	92.13	**56.18**	85.39	77.53	83.15
Wine-SRC	64.04	93.26	**96.63**	96.63	77.53	80.90	**92.13**	**96.63**	46.07	88.76	88.76	**94.38**
Glass-SVM	55.14	76.64	80.37	81.31	34.58	32.71	51.40	47.66	70.09	62.62	76.64	75.70
Glass-KNN	**56.07**	77.57	77.57	75.70	**39.25**	38.32	48.60	49.53	**73.83**	71.96	75.70	76.64
Glass-SRC	54.21	**91.59**	**97.20**	**95.33**	36.45	**44.86**	**59.81**	**70.09**	70.09	**79.44**	**95.33**	**95.33**
Libras Movement-SVM	63.33	53.89	51.67	47.22	61.11	61.67	65.56	65.00	51.67	53.33	61.67	63.89
Libras Movement-KNN	59.44	43.33	33.89	31.11	58.33	60.00	58.89	59.44	43.33	55.00	54.44	56.11
Libras Movement-SRC	**82.78**	**83.33**	**83.33**	**83.33**	**74.44**	**80.56**	**82.78**	**84.44**	**67.22**	**73.89**	**81.67**	**84.44**
Wine Quality-SVM	46.25	48.63	55.88	56.50	37.63	44.75	**53.88**	54.75	**55.25**	**59.13**	**58.13**	**58.13**
Wine Quality-KNN	46.38	48.63	52.88	53.00	43.13	46.25	52.25	51.25	54.50	55.88	50.25	52.75
Wine Quality-SRC	**46.63**	**50.75**	**56.25**	**58.63**	**46.38**	**50.75**	52.88	**56.00**	48.00	55.50	54.50	56.88
Anneal-SVM	76.35	76.35	76.35	76.35	76.35	76.35	76.35	76.35	76.35	76.35	76.35	76.35
Anneal-KNN	76.11	76.35	77.34	79.56	44.83	**82.02**	**83.74**	**83.74**	81.28	82.27	**85.96**	79.06
Anneal-SRC	**78.57**	**79.56**	**79.80**	**80.79**	**76.60**	80.30	79.80	79.31	**81.28**	**84.48**	79.06	**80.54**

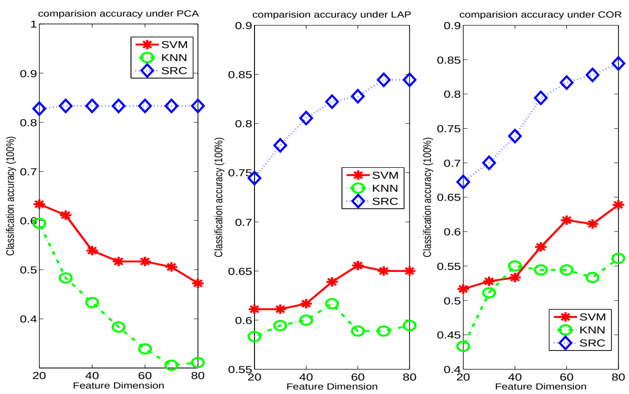

Figure 1. Classification results for data Libras Movement

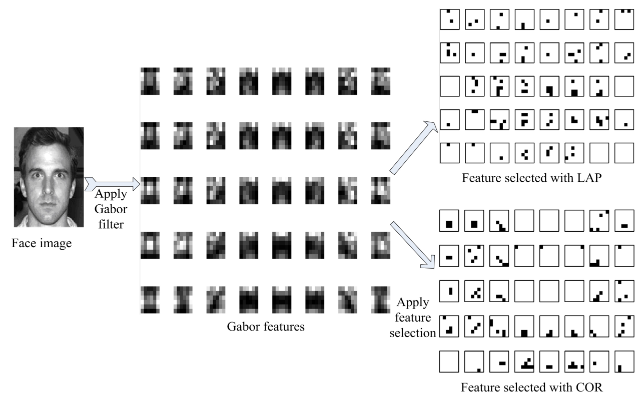

Figure 2. Face image process

40%, 60% and 80% of total feature size) with different dimension reduction methods (PCA, LAP and COR). The highest accuracies are highlighted among SVM, KNN and SRC. In most cases, the results of SRC have higher accuracies compared with SVM and KNN.

3.2. Experiments on Face Recognition

The face dataset is from the extended Yale face database B. There are 2414 faces belonging to 38 people, captured with various lighting conditions. Each face image has a size of 54×48 pixels. Gabor features have been frequently used in image analysis. A recent work (Yang & Zhang, 2010) using Gabor features for face recognition has reported very promising performance. This motivated us to investigate SRC-SPDR performance on Gabor features. First we extract Gabor features for each image with the parameters same as in (Yang & Zhang, 2010). A set of Gabor filters, which contains 5 scale levels and 8 orientations, are applied to each face image. There are totally 40 Gabor filters, and each Gaborface is with the size of 6×6, as shown in the middle of Figure 2. We then apply feature selection methods, i.e. PCA, LAP and COR, based on Gabor features of 1440 dimension.

On the right side of Figure 2, we show an example of selected 100 features by LAP and COR methods respectively. From this figure, it is easy to tell that the selected features by two methods are very different. However, the performance of LAP features and COR features are similar with 3 classifiers, as seen in Figure 3. Hence we may conclude that, there is much redundancy in Gabor features, and structure preserving dimension reduction is meaningful in this case. In Figure 3, with feature selection methods LAP and COR, SRC shows clear performance advantage. Although the performance of SRC with PCA is a little less than SVM with PCA, but the accuracy is still greater than 92% and remaining stable at 95% when the feature dimension is greater than 300. Overall, the accuracy of SRC is competitive with SVM and KNN.

Table 3. Comparisons of classification accuracy(%) based on low dimensional Gabor features

DIMENSION(D)	10	20	30	40	50	60	70	80	90	100
PCA-SVM	27.81	69.62	83.03	87.42	89.65	90.98	91.06	91.80	**92.88**	**94.12**
PCA-KNN	28.81	63.82	75.17	78.39	83.03	85.60	85.84	88.41	89.07	90.07
PCA-SRC	**36.42**	**73.84**	**84.35**	**88.49**	**90.23**	**91.23**	**91.31**	**92.05**	92.30	92.30
LAP-SVM	22.43	44.95	56.87	66.56	68.29	71.94	75.33	78.23	79.88	80.88
LAP-KNN	26.32	44.87	54.22	58.03	61.67	62.67	65.98	67.05	69.12	69.04
LAP-SRC	**38.74**	**66.47**	**77.07**	**82.04**	**83.69**	**85.68**	**86.09**	**88.08**	**88.49**	**88.49**
COR-SVM	27.81	46.03	60.43	69.04	70.86	74.92	77.24	78.73	78.56	80.46
COR-KNN	36.09	48.51	57.20	61.75	62.67	64.57	67.63	69.04	68.54	70.78
COR-SRC	**42.38**	**64.49**	**77.40**	**83.36**	**85.76**	**88.00**	**89.32**	**89.90**	**89.74**	**90.65**

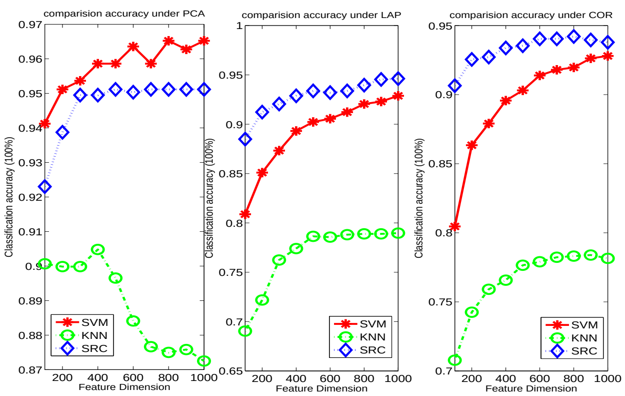

Figure 3. Classification results for high dimensional Gabor face features

Figure 3 shows the recognition rates over a wide dimension range, from 100 to 1000. It is particularly interesting to study the classification performance of structure preserving at very low dimensions. In Table 3, classification performance for lower dimension of Gabor features are shown. The dimension varies from 10 to 100, out of the original size of 1440. In the COR and LAP cases, we can observe that SRC results always achieve higher accuracy than SVM and KNN. With PCA, the SRC results are the highest when the dimension is smaller than 80.

4. Conclusion

This paper presented a comprehensive study on various structure preserving dimension reduction (SPDR) techniques within a sparse representation based classification (SRC) framework. The integrated SRC with SPDR methods were tested on both UCI feature space datasets and nature face image dataset. Experimental results demonstrated the effectiveness of SRC-SPDR framework. In our future work, more experiments may be needed to evaluated the robust of these methods on more challenging data sets, and relative theoretical analysis is needed for justification of the proposed framework.

References

Frank, A. and Asuncion, A. UCI machine learning repository, 2010. URL `http://archive.ics.uci.edu/ml`.

Guyon, I. and Elisseeff, A. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

He, X., Cai, D., and Niyogi, P. Laplacian score for feature selection. In *Neural Information Processing Systems*, 2005.

Kim, S., Koh, K., Lustig, M., Boyd, S., and Gorinevsky, D. An interior-point method for large-scale l1-regularized least squares. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):606 –617, 2007.

Lacoste-Julien, S., Sha, F., and Jordan, M. I. DiscLDA: Discriminative learning for dimensionality reduction and classification. In *Neural Information Processing Systems*, 2008.

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. Online Dictionary Learning for Sparse Coding. In *International Conference on Machine Learning*, 2009.

Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1996.

Witten, I. H. and Frank, E. Data mining: practical machine learning tools and techniques with Java implementations. *ACM SIGMOD Record*, 31, 2002.

Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., and Ma, Y. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210 –227, 2009.

Yang, M. and Zhang, L. Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. In *ECCV (6)*, pp. 448–461, 2010.