PROCEEDINGS OF SCIENCE

Relationship between the number of labeled samples and classification accuracy based on sparse representation

Qianqi Zhou[1]
School of Computer Science and Engineering, Beifang University for Nationalities,Yinchuan, 750021,China

Chunmei Zhang [2]
School of Computer Science and Engineering,Beifang University for Nationalities,Yinchuan, 750021,China
E-mail: chunmei66@hotmail.com

Yunbin Zhang[3]
School of Computer Science and Engineering, Beifang University for Nationalities, Yinchuan, 750021,China

Sparse decomposition algorithm can be implemented to classify high dimensional data without dimension reduction. Empirical examinations have shown that the classification accuracy of the hyperspectral data can be significantly improved based on sparse representation framework. However, the classification accuracy of hyperspectral data is reduced greatly if the quantity of samples is too small. This paper represents an attempt to fill the gap represented by the absence of the lower bound of labeled sample quantity of sparse representations in supervised classification task, in order to focus future analyses towards the most relevant aspects of image classification accuracy and data labeling cost. Based on the boundary conditions of the unique solution of sparse representation, the relationship between the classification accuracy of hyperspectral data and the quantity of labeled samples is analyzed theoretically in this paper. The correlation curves of the sample quantity and the classification accuracy is figured with the Kennedy Space Center (KSC) hyperspectral data.

CENet2015
12-13 September 2015
Shanghai, China

[1]Speaker
[2]Corresponding Author
[3]This work is supported by National Natural Science Foundation of China (61461002) and Innovation Projects for graduate students of Beifang University for Nationalities.

PoS(CENet2015)014

1. Introduction

The paradox of high dimension of hyperspectral data is that, in one hand, many classification methods are not appropriate for dealing with high dimensional data, so dimension reduction becomes a main approach. On the other hand, high dimension means high information. If dimension reduction methods are implemented, the valuable information will be lost.

Many classification algorithms have been implemented to classify hyperspectral data. Support Vector Machine (SVM) is a classical algorithm, but it is sensitive to the kernel function. HSVM (Hierarchical SVM) is the improvement of the SVM which combine the Max-cut tree and SVM together in order to enhance the classification accuracy [1]. Dictionary learning of sparse representation can obtain better classification accuracy [2]. However, dictionary learning is time consuming.

Sparse decomposition algorithm is a fast and effective framework and can be implemented to classify high dimensional data. The framework has the advantage that it does not need dimension reduction, model selection and dictionary training. Each test sample can be sparsely represented with only a few non-zero decomposition coefficients by a few training samples of an over-complete dictionary which are constructed by training samples directly. According to the approximation error, test samples can be categorized to the corresponding classes. Several researchers have improved the classification accuracy of hyperspectral data based on sparse representation framework [3-5].

Empirical examinations have shown that the accuracy of hyperspectral classification is reduced greatly if the quantity of labeled samples is too small. However, identifying more labeled samples will make the cost rise. To get the lower bound of the sample quantity and analyze the relationship between the quantity of labeled samples and classification accuracy, we deduce a formula which is the relationship between the residual error of sparse representation and the number of dictionary atoms. Then through empirical data, the correlation curve between the number of labeled samples and classification accuracy is figured.

The rest of paper is organized as follows. In the second section, we introduce the theoretical analysis of the relationship between the number of labeled samples and classification accuracy. In the third section, we analyze and show the experimental results. Finally, the fourth section concludes this paper.

2.Theoretical Analysis

In this section, we introduce the theoretical analysis of the relationship between the number of labeled samples and classification accuracy.

2.1 Sparse Representation Algorithm

We can construct a set of dictionaries according to different categories. Each dictionary gets the labeled samples of the same category from hyperspectral data to form the over-complete dictionary. And let $D_c=[d_1, d_2,..., d_M](D \in R^{N*M})$ denotes the dictionary of *cth* category.

According to the theory of sparse representation, each test sample can be sparsely represented by a few training samples of an over-complete dictionary with only a few non-zero decomposition coefficients. We can get sparse decomposition coefficient according to formula (2.1).

$$\hat{\alpha}(f) = \arg\min_{\alpha} \|f - D_c\alpha\|_2^2 + \lambda\|\alpha\|_1 \tag{2.1}$$

The above formula is an l_1-regularized least squares problem. Where $\alpha = [\alpha_1, \alpha_2, \cdots \alpha_M]^T$ is the coefficient of sparse representation. Regularization parameter λ is used to balance the data reconstruction error and sparsity. The threshold of parameter λ is selected according to empirical value.

Then the test sample f is categorized to the corresponding classes according to formula (2.2).

$$\min_c R_c(f) = \min_c \left(\|f - D_c \hat{\alpha}\|_2 \right) \tag{2.2}$$

Where $R_c(f)$ is the residual to reconstruct f by training samples in the category c. The smaller the value $R_c(f)$ is, the higher the possibility of f belongs to the category of *cth*.

2.2 Coherent

The coherent coefficient μ is a character of dictionary. Its value equals to the maximum of inner product between two distinct atoms in the dictionary [6].

$$\mu := max_{i \neq j} |\langle d_i, d_j \rangle| \qquad s.t.\ \ i, j \in [1, \dots ,M] \tag{2.3}$$

Generally, $0< \mu <1$. If μ has a very small value, we say that the redundant dictionary is incoherent, and obviously, the coherent coefficient of an orthogonal basis equals zero. When μ=1, then the dictionary contains at least two identical atoms [7]. Furthermore, in finite dimension space of N, the coherence of a concatenate dictionary of two orthonormal bases is $\mu >= 1/\sqrt{N}$ [6]. For general redundant dictionary, a lower bound of coherence coefficient is [8]

$$\mu >= \sqrt{\frac{M-N}{N(M-1)}} \tag{2.4}$$

Where M is the number of dictionary atoms, N is the dimensionality of data sample.

2.3 Recovery and Convergence of MP

Basis pursuit (BP),Matching pursuit (MP) and its variants are the classic algorithm of sparse representation. For the Exact Recovery Condition of BP and OMP (Orthogonal matching pursuit), J.A.Tropp has given a unified sufficient condition [4]:

$$m < \frac{1}{2}\left(1 + \mu^{-1}\right) \tag{2.5}$$

As long as m which is the number of nonzero entries of the coefficients of sparse representation satisfies above formula (2.5), OMP and BP both recover every superposition of m atoms from D_c .

In most cases, the signal or image can only be expressed by linear combination of some atoms in the actual dictionary. For MP algorithm, the approximation performance can be described as follows [9]:

Theorem: Let $\{f_n\}$ be a sequence of approximants to $f \in H$ where d_{in} is the corresponding atoms selected through MP. Let

$$m < \frac{1}{4}\left(1 + \mu^{-1}\right) \tag{2.6}$$

and $\tilde{f}_m = \sum_{i \in \tilde{I}m} \alpha_i d_i$ be a best m approximant to f from D , i.e.

$$\|\tilde{f}_m - f\| = \sigma_m(f) := \inf \{ \|f - D_I \alpha\|, card\,(I) <= m\ , \alpha \in C^I \} \tag{2.7}$$

then, there is an number N_m such that

the error after N_m steps satisfies

$$\|f_{N_m} - f\| <= \sqrt{1+4m}\,\sigma_m \tag{2.8}$$

during the first N_m steps MP picks up atoms from the best m-term approximant: $i_n \in \tilde{I}_m$;
if $\sigma_m^2 < 3\sigma_1^2 / m$, then

$$N_m <= 2 + m.\frac{4}{3}.\ln\frac{3\sigma_1^2}{m\sigma_m^2} \tag{2.9}$$

2.4 Monotonicity Analysis of Residual Error

From the formula (2.6) and (2.8), we can get a new inequality about the relationship between $\|f_{N_m} - f\|$ and coherent coefficient μ.

$$\|f_{N_m} - f\| < \sqrt{2+\mu^{-1}} \tag{2.10}$$

From the formula (2.4) and (2.10), we can derive the following inequality between the number of dictionary atoms and the residual error of sparse approximate:

$$\|f_{N_m} - f\| < \sqrt{2+\sqrt{\frac{N(M-1)}{M-N}}}\sigma_m \tag{2.11}$$

The dimension of dictionary is unchanged, so, in inequality (2.11), N can be considered as a constant.

Then inequality (2.11) can be written as

$$\|f_{N_m} - f\| < \sqrt{2+\sqrt{N+\frac{N(N-1)}{M-N}}}\sigma_m \tag{2.12}$$

So, there exist an $\lambda_m \in$ (0,1]. $\|f_{N_m} - f\|$ will decrease with the increase of independent variable M.

$$\|f_{N_m} - f\| = \lambda_m\sqrt{2+\sqrt{N+\frac{N(N-1)}{M-N}}}\sigma_m \tag{2.13}$$

By analyzing the monotonicity of equality (2.13), we can find that variable $\|f_{N_m} - f\|$ will decrease with the increase of independent variable M. The classification criterion of sparse representation method is measured by $\|f_{N_m} - f\|$. In another word, the smaller the residual error $\|f_{N_m} - f\|$ is, the higher the classification accuracy is. When the number of atoms of dictionary increases, the residual error will decreases gradually, and the classification accuracy will rise, so classification accuracy is positively correlated with the number of atoms. Because the atoms of dictionary are labeled training samples, so the classification accuracy is positively correlated with the number of the labeled samples.

3. Experimental Analysis and Conclusion

KSC hyperspectral data is used in our experiment. The results of our experiment advocate that the number of labeled samples is positively correlated with classification accuracy. The number of labeled samples plays a critical role on classified accuracy.

3.1 Data set

The specific information of hyperspectral data that employed in experiment includes the following.

The hyperspectral image data was acquired by the National Aeronautics and Space Administration (NASA) Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) instrument. AVIRIS acquired image data over the Kennedy Space Center (KSC), Florida, on March 23, 1996. The data has 224 bands and the center wavelengths is from 400~2500nm. After removing water absorption and low signal-to-noise (SNR) bands, 176 bands were selected for the analysis. The hyperspectral data contains 13 typical classes of land cover and 5211 sample points. The detail of the data set is tabulated in table 1.

Class	Class name	No.samples
1	Scrub	761(14.6%)
2	Willow swamp	243(4.66%)
3	Cabbage palm hammnck	256(4.92%)
4	Cabbage palm/oak hammnck	252(4.84%)
5	Slash pine	161(3.07%)
6	Oak/broadleaf hammock	229(4.38%)
7	Hardwood swamp	105(2.0%)
8	Graminoid marsh	431(8.27%)
9	Spartina marsh	520(9.99%)
10	Cattail marsh	404(7.76%)
11	Salt marsh	419(8.04%)
12	Mud flats	503(9.66%)
13	Water	927(17.8%)

Table 1: The Information of KSC Data

3.2 Data Preprocessing and the Algorithm Steps

In the experiment, firstly, 25% data samples are selected randomly into the test set, the rest of 75% data are used as the pool of training examples. Then nine different training sets with different proportion of samples selected randomly from the training example pool are investigated, e.g. 5%, 8%, 10%, 12%, 15%, 20%, 30%， 50% and 75%. Each training sets with corresponding proportion are executed 10 times. Pixel values with respect to each band are normalized. The experimental steps are as follows:

Classification algorithm
Input: X: Training set (labeled example set) F: Test set (Unlabeled example set) Output: $c(y)$: the category of each test sample f. Each training sample $x \in X$ or test sample $f \in F$ is normalized. Use the training set to form the over-complete dictionary D_c. Use lasso algorithm to decompose the test sample sparsely to get the sparse representation coefficient $\hat{\alpha}$. According to formula (2.2)， calculate the residual error $R_c(f)$ of test sample which was sparse approximation by dictionary composed of samples of the *cth* category. Then the test sample is categorized to the corresponding category according to the minimum residual error. Stop repeat step 3and 4 until all the test samples are categorized.

3.3 Experiment Results and Analysis

The classification accuracy is the average of the ten experimental results. The classification method used for comparison is HSVM. Table 2 compares the classification performances of HSVM and sparse representation (SR) at diffident sampling rates. For the two compared algorithms on each data set, the average classification accuracy increase as the size of labeled data set gets large and SR consistently demonstrates better performance compared to HSVM.

Training rate	Number of sampling	HSVM	SR
5%	256	85.562	86.603
8%	411	86.042	88.753
10%	516	86.779	89.434
12%	620	86.901	90.245
15%	776	87.192	91.186
20%	1037	88.329	91.867
30%	1557	89.265	93.145
50%	2601	90.252	94.147
75%	3904	91.889	94.889

Table 2: The Classification Accuracies of HSVM and SR

Furthermore, table 2 describes the corresponding relationship between classification accuracy and different rate of the training sample, namely, the number of training sample. Figure 1 draws the relation curve between classification accuracy and the number of training sample according to the specific experiment results.

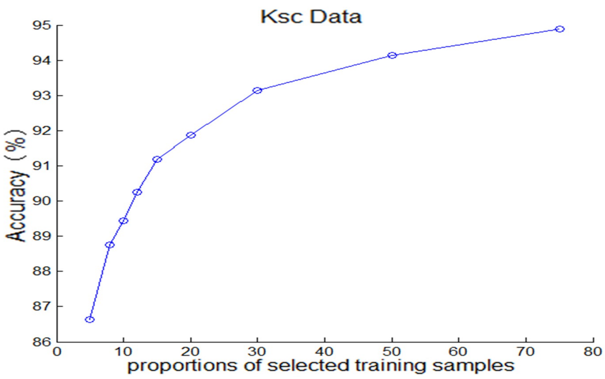

Figure 1: the relation curve between classification accuracy and the number of training sample

Through analyzing table 2 and figure 1, we can easily find that the classification accuracy monotonically increase with the number of labeled samples. In the experiment of SR, the proportions of selected training samples are gradually increased from 5% to 75%. The classification accuracy is gradually increased from 86.603% to 94.889%.

4. Conclusion

In this paper, we proved that classification accuracy of hyperspectral data is positively correlated with the number of labeled samples. According to the theory of sparse representation, firstly, we derive the relationship between the number of labeled samples and classification accuracy. Then through the experiment, we give the correlation curves between the number of labeled samples and classification accuracy. When the number of labeled samples reaches 10%

PoS(CENet2015)014

or more, the classified accuracy can be ensured more than 90%. Obviously, the number of labeled samples plays a critical role on classification accuracy.

References

[1] Y.C. Chen, M. M. Crawford, J. Ghosh. *Integrating support vector machines in a hierarchical output space decomposition framework*[C]. Geoscience and Remote Sensing Symposium, IEEE International. Piscataway,N.J. pp,949-952(2004)

[2] J. Mairal, F. Bach, J. Pronce, G. Sapiro. *Online Learning for Matrix Factorization and Sparse Coding*[J]. Journal of Machine Learning Research.11(1):19-60(2010)

[3] M.S. Cui, S. Prasad. *Class-Dependent Sparse Representation Classifier for Robust Hyperspectral Image Classification*[J]. Geoscience and Remote Sensing, IEEE Transactions on.53(5): 2683-2695(2015)

[4] Y.F. Tang, X.M. Li, Y. Xu, Y. Liu, J.Z. Wang, C.Y. Liu . *Hyperspectral image classification using sparse representation-based classifier*[C]. Geoscience and Remote Sensing Symposium,IEEE International. Piscataway,N.J. pp,3450-3453(2014)

[5] S. Qazi, L.X. Shi, L.M. Tao, S.Q. Yang. *A l1-minimization Based Approach for Hyperspectral Data Classification*[J]. Key Engineering Materials.500: 675-681 (2012)

[6] J.A. Tropp. *Greed is good: algorithmic results for sparse approximation*[J]. Information Theory,IEEE Transactions on.50(10):2231-2242 (2004)

[7] C.M. Zhang, Z.K. Yin, X.D. Chen, M.X. Xiao. *Signal overcomplete representation and sparse decomposition based on redundant dictionaries*[J]. Chinese Science Bulletin. 50(23): 2672-2677(2005)

[8] A.F. Abdelnour, I.W. Selesnick. *Symmetric nearly shift-invariant tight frame wavelets*[J]. Signal Processing, IEEE Transactions on. 53(1): 231-239(2005)

[9] R. Gribonval, P. Vandergheynst. *On the exponential convergence of Matching Pursuits in quasi-incoherent dictionaries*[J]. Information Theory ,IEEE Transactions on.52(1):255-261(2006)