

A Novel Overcomplete Dictionary Training Based on Empirical Mode Decomposition and Its Performance Analysis

Shikui Wang*, Yufeng Shao

College of Electronic and Information Engineering, Chongqing Three Gorges University

Wanzhou District of Chongqing City, China

**wangshikui2002@aliyun.com*

Abstract.

In this paper, a novel overcomplete dictionary training method which is based on empirical mode decomposition is presented. The IMFs by empirical mode decomposition take part in the training of overcomplete dictionary, and K-SVD algorithm is adopted in the training process. Simulation results show that, compared with the dictionary trained directly from the original speech signals, the overcomplete dictionary has sparser representation for the speech signals, and thus has higher reconstructed speech quality.

Keywords: overcomplete dictionary; EMD; sparse representation; K-SVD algorithm; DCT

1. Introduction

In recent years, overcomplete dictionary learning are absorbing more and more attention in signal processing [1- 4]. Michael Aharon et al, presented K-SVD algorithm for designing of overcomplete dictionary [5]. Ivana Tomic and Pascal Frossard [6] summarized the methods of dictionary learning and pointed out that understanding the underlying causes of signals or the relevant information in observations is a challenging task when the training samples are imperfect.

In this paper, we propose an overcomplete dictionary training method based on empirical mode decomposition (EMD) [7]. The decomposed IMFs are taken as the training signals. The quality of the reconstructed speech through the trained

overcomplete dictionary is compared with overcomplete dictionary trained by original speech signals. Simulation results indicate that, the speech signals have sparser representation under the trained overcomplete dictionary based on EMD.

2. Overcomplete Dictionary Training Based on EMD

Our goal is to train an overcomplete dictionary matrix $\mathbf{D} \in \mathbf{R}^{n \times K}$ which contains K prototype signal-atoms for columns $\{\mathbf{d}_j\}_{j=1}^K$. For any signal $\mathbf{y} \in \mathbf{R}^n$, it can be represented as a sparse linear combination of the signal-atoms. We can represent the signal \mathbf{y} exactly as $\mathbf{y} = \mathbf{D}\mathbf{x}$, or approximately as $\mathbf{y} \approx \mathbf{D}\mathbf{x}$ which satisfies $\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_p \leq \varepsilon$, with $\|\cdot\|_p$ indicating the p -norm of a vector. In this paper, we set $p = 2$. The vector $\mathbf{x} \in \mathbf{R}^K$ contains the representation coefficients of the signal \mathbf{y} . When $n < K$ and \mathbf{D} is a full-rank matrix, \mathbf{D} is a typical overcomplete dictionary. Given \mathbf{y} and \mathbf{D} , an infinite number of solutions of $\mathbf{y} = \mathbf{D}\mathbf{x}$ are available, so we must set some constraints on the solution. Generally, the solution with the fewest number of nonzero coefficients is the most significant, which is also called as the sparsest solution. We will determine the overcomplete dictionary matrix \mathbf{D} that yields the sparsest representation for the training signals. Unlike all the other training methods for overcomplete dictionary in which the original speech signals are used as the training signals, we will at first decompose the speech signals into intrinsic mode functions (IMFs), and then adopt the IMFs as the training signals. By using K-SVD algorithm, the overcomplete dictionary is trained. In the testing stage, an easy, efficient orthogonal matching pursuit (OMP) algorithm is adopted to seek the sparse solutions.

2.1 Empirical mode decomposition

Empirical mode decomposition is presented mainly to analyze nonlinear, non-stationary signals. Two steps are taken in Huang's theory [7]: the first step is

to preprocess the data by the empirical mode decomposition method; the second step is to apply the Hilbert transform to the decomposed intrinsic mode functions (IMFs). In this paper, we only need the first step and decompose the speech signals into IMFs. The decomposed IMFs are taken as the basic training signals. The reason why we use the IMFs as the basic training signals is that the decomposed IMFs through EMD are more fundamental and able to capture the causes underlying the speech signals which are produced by human vocal organs. Fig. 1 shows one segments of voiced speech with 148 sampling points and its IMFs. According to the property of speech, one segment of speech can decompose into 3 to 9 IMFs. Usually, voiced speech has fewer IMFs, and unvoiced speech has more IMFs. Special attention must be paid to the end effect in EMD [8].

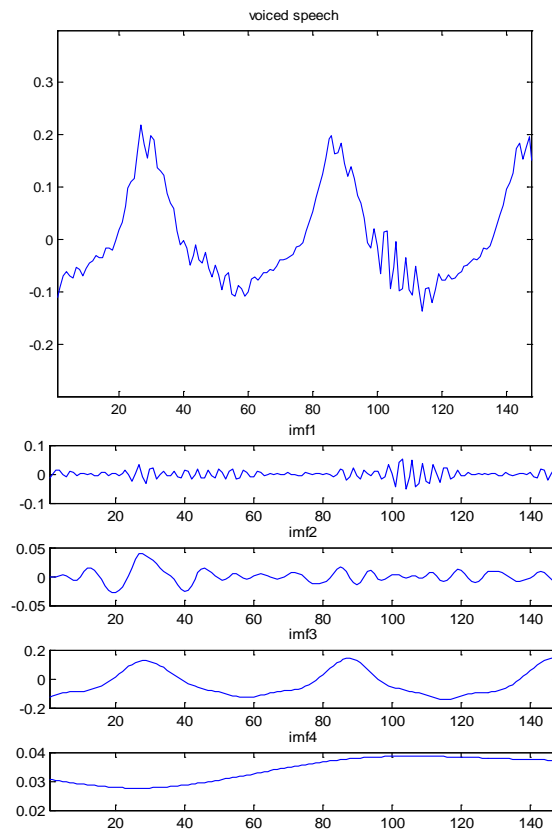


Figure 1. one segment of voiced speech and its IMFs

2.2 K-SVD algorithm for overcomplete dictionary training

K-SVD algorithm [5] was proposed for training overcomplete in order to achieve sparse signal representations. K-SVD is an iterative method that alternates between sparse coding of the samples based on the current dictionary, and a process of updating the dictionary atoms to better fit the data. In this paper, K-SVD algorithm uses the IMFs by EMD instead of original speech signals to train the overcomplete dictionary. Two other overcomplete dictionaries, that is, fixed DCT dictionary and the dictionary trained by the original speech signals, will compare with the dictionary trained in this paper in the aspects of sparsity and reconstructed speech quality.

2.3 Orthogonal matching pursuit algorithm

In 1993, Pati et al [9] proposed a modification to Matching Pursuit algorithm, and referred their algorithm as Orthogonal Matching Pursuit (OMP). After that, OMP has been intensely researched and applied [10, 11]. Compared to globally-optimized sparse representation algorithm, such as Basis Pursuit (BP), OMP has the merits of faster convergence and easier realization, and so it is widely adopted in many algorithms. In this paper, we will use OMP to seek the sparse representation of a signal under some overcomplete dictionary.

3. Simulation Designing and Results

3.1 Simulation designing

In our simulation, 20 segments of Chinese speech taken from speech database are utilized to train the overcomplete dictionary. Before training, all the speech segments are divided into signal vectors each consists of 148 sampling points, with 20 overlapping points for eliminating the end effect of EMD. Fig. 2 shows the segmentation of speech samplings. The EMD signals are the parts that are processed by EMD. The overlapping is the part that two consecutive EMD signals overlap and it is designed to eliminate the end effect of EMD. After the EMD

signals are decomposed into intrinsic mode functions (IMFs), the corresponding overlapping signals of each IMFs are discarded and the remained IMF training signals are taken as the columns of the training data matrix.

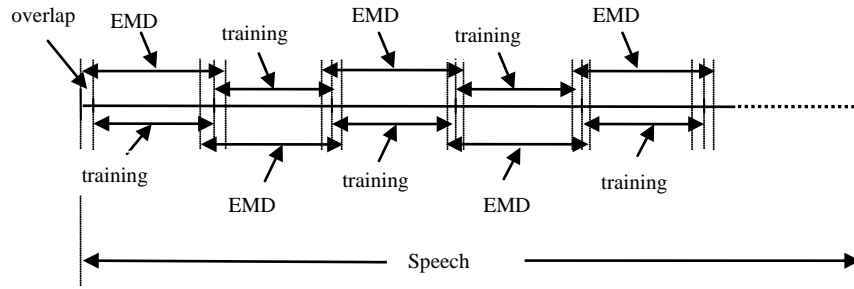


Figure 2. Speech signal segmentation for EMD

To train the overcomplete dictionary, 20 segments of speech signals from Chinese speech database are used, and each of about 1-2 seconds. Considering the length of any significant phoneme, the length of training signals is 128 sampling points. The number of atoms in the overcomplete dictionary is 1280. To simplify the training process, we only select 3840 atoms in front of the training matrix as the training signals. Among the 3840 atoms, 1280 atoms form the initial dictionary, and the remained 2560 atoms form the data for updating dictionary, depicted as in Fig. 3.

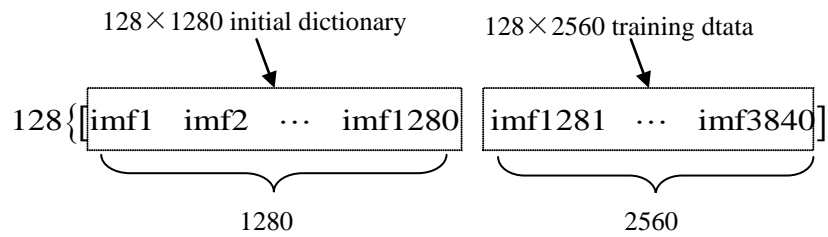


Figure 3. Structure of training signals

K-SVD algorithm is used to train the overcomplete dictionary. The CPU of PC is Pentium R 3.4Hz.

3.2 Simulation results

About two hours are spent in the training of overcomplete dictionary in this simulation. When the training is ended, we obtain the dictionary $\mathbf{D}_{128 \times 1280}$, with

1280 atoms, and any signal y of length of 128 can be sparsely represented as $y = \mathbf{D}x$ with x being a sparse 128×1 vector which has very few nonzero entries compared to its dimension. We will take two steps to compare the performance of the newly-trained overcomplete dictionary.

First, the dictionary \mathbf{D} is compared with the dictionary \mathbf{D}' which is trained directly through the original speech signal, that is, for the same testing signal y , the sparsity of x and x' will be comparatively analyzed in $y = \mathbf{D}x$ and $y = \mathbf{D}'x'$. This can be done through randomly selecting some testing signals outside the training data set. OMP is adopted to seek the sparse solution in $y = \mathbf{D}x$ and $y = \mathbf{D}'x'$. Table 1 shows the sparsity under different energy error.

Table 1. Comparison between two trained overcomplete dictionary

Energy error rate	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
Number of nonzero atoms (\mathbf{D})	7.8	25.	53.	75.	91.2
		8	1	4	
Number of nonzero atoms (\mathbf{D}')	15.	35.	63.	90.	107.
	2	4	6	8	1

Second, the overcomplete dictionary \mathbf{D} is compared with 10 times overcomplete DCT dictionary. The number of basic atoms in overcomplete DCT dictionary are also the same as that of \mathbf{D} . When some fixed rate atoms are selected to reconstruct the speech, subjective listening tests indicate that the reconstructed speech through dictionary \mathbf{D} has higher mean opinion score (MOS).

4. Summary

In this paper, we propose an overcomplete dictionary training project based on empirical mode decomposition. Simulation shows that the speech has sparser representation under the newly-designed dictionary, and thus has higher reconstructed speech quality than that of the other overcomplete dictionary. In the near future, we will explore the nature of intrinsic mode function and make use of it in the application of speech enhancement, emotion recognition and so on.

Acknowledgements

This work was supported by 2013 Chongqing University Innovation Team Founding (No. KJTD201320) and Chongqing Educational Committee Science and Technology Research Item(KJ131116).

References

- [1] Rusu C, Dumitrescu B. An initialization strategy for the dictionary learning problem. ICASSP2014, Page(s): 6731 - 6735
- [2] Khalid M U, Seghouane A K. A single SVD sparse dictionary learning algorithm for FMRI data analysis. 2014 IEEE Workshop on Statistical Signal Processing (SSP), Page(s): 65 – 68
- [3] Zhenni Li, Shuxue Ding, Yujie Li, et al. Improving dictionary learning using the Itakura-Saito divergence. 2014 IEEE China Summit & International Conference on Signal and Information Processing. Page(s): 733 - 737
- [4] Tillmann A M. On the Computational Intractability of Exact and Approximate Dictionary Learning. IEEE Signal Processing Letters, vol.22 (1), Page(s): 45 - 49
- [5] Aharon M, Elad M, Bruckstein A. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. IEEE Transactions on Signal Processing, vol.54 (11), Page(s): 4311 - 4322
- [6] Ivana Tomic, Pascal Frossard. Dictionary Learning. IEEE Signal Processing Magazine, vol.28(2), Page(s): 27 - 38
- [7] N. E. Huang et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol.454, Page(s): 903–995
- [8] Haiyan Quan, Zengli Liu, Xinling Shi. A new processing method for the end effect problem of empirical mode decomposition. 2010 3rd International Congress on Image and Signal Processing (CISP), vol.7, Page(s): 3391 - 3394

- [9] Pati Y C, Rezaifar R, Krishnaprasad P S. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. ACSSC1993, vol.1, Page(s): 40 - 44
- [10] Tropp J A, Gilbert A C. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. IEEE Transactions on Information Theory, 2007, vol.53 (12), Page(s): 4655 - 4666
- [11] Jie Chen, Huo X. Theoretical Results on Sparse Representations of Multiple-Measurement Vectors. IEEE Transactions on Signal Processing, 2006, vol.54 (12), Page(s): 4634 – 4643