DOI: 10.6084/m9.figshare.9788426

Yan Yang
Master of Engineering, Research Direction: Electronic Communications
Yancheng industrial vocational technical college school of automotive engineering, Yancheng Jiangsu 224005, China
Li Zhao
School of Information Science and Engineering of Southeast University
Nanjing 210096, China

A STUDY ON THE SPEAKER RECOGNITION ALGORITHM BASED ON BAYESIAN COMPRESSED SENSING THEORY

Abstract. *Speaker recognition technology is widely used in the Internet and communication filed. In recent years, compressed sensing theory attracted wide attention in and abroad. Having broken through the limitations of the Nyquist sampling rate, it can sampled compressible signals while compressing them at the same time.As a new theory, compressed sensing theory was brought into the filed of speaker recognition technology where huge breakthrough is in great demand to bring hope to enhance the performance of speaker recognition system. Aiming at the text independent speaker recognition technology, this paper made a profound study on the Bayesian compressed sensing algorithm. According to the characteristics of the sparse coefficient in the algorithm, a Gaussian prior assumption is introduced. And then a speaker recognition algorithm based on Bayesian compressed sensing is proposed.*

Keywords: ***speaker; recognition; compressed sensing; sparse representation; BCS GMM***

Introduction

Automatic speaker recognition, a technology that automatically identifies the speakers by analyzing the speakers' voices , is widely used in the field of communication and Internet. Since the concept of "voiceprint" has been put forward by L.G. Kesta of Bell Laboratory , scholars have carried out a lot of research on different modules of speaker recognition technology, and gained a lot. At present, the research of text-related speaker recognition technology has been relatively mature with the highrecognition rate, but the performance of text-independent speaker recognition technology still has a good space to promote. In addition, in view of the fact that most of the speech signals in the application are noisy, it is also crucial to find a speaker recognition algorithm with better anti-noise performance. In 2004, the compressed sensing theory proposed by Donoho et al. could acquire the signal at a rate lower than required by Nyquist sampling theorem.

The advent of compressed sensing theory is benefit for the improvement of speaker recognition technology performance. Compressed sensing theory mainly includes three aspects: sparse representation of signals, design of observation matrix and signal reconstruction algorithm. Among them, signal reconstruction is of great concern. Candes et al proved the signal reconstruction problem which is to solve the minimum 0-norm problem [1], but this problem is a NP-Hard problem. Aiming at the extremely complex computation of the algorithm, then many suboptimal equivalent solutions have been proposed, mainly including the minimum 1-norm algorithm, matching pursuit algorithm and so on. In 2008, Shihao Ji proposed an algorithm to solve the problem of compressed sensing reconstruction in Bayesian framework [2], thus putting forward a new idea for the solution of this problem. Compared with the traditional BP algorithm and the OMP algorithm for the sparse signal, the solution obtained by the Bayesian reconstruction algorithm based on relevance vector machine is more sparse and closer to the norm estimation. On the other hand, sparse bayesian learning, the core of bayesian compressed sensing, has been proved to be successfully applied in the field of pattern classification [3-4]. In this paper, Bayesian compression sensing algorithm and speaker recognition technology are combined in order to achieve better recognition results.

Bayesian compression sensing algorithm

Firstly, it is assumed that the signal x is compressible under the matrix Ψ , when the sparse coefficient is α . α_s is the largest M elements of vector α , similarly α_e consist of the smallest $N-M$ elements of vector α .

$$\alpha = \alpha_s + \alpha_e \tag{1}$$

$$y = \Phi\alpha = \Phi\alpha_s + \Phi\alpha_e = \Phi\alpha_s + n_e \tag{2}$$

$$n_e = \Phi\alpha_e \tag{3}$$

y is the observation vector. Φ is the CS matrix, which is the product of the observability matrix and the sparse matrix. According to the central limit theorem, the elements in n_e should be approximately subject to the gaussian distribution with the mean value of 0. It is also noted that the observation vector of CS may be noisy, which is represented by m. Therefore

$$y = \Phi\alpha_s + n_e + n_m = \Phi\alpha_s + n \qquad (4)$$

n obeys the normal distribution of the mean value of 0 and the unknown variance, thus Gauss likelihood function is obtained

$$p(y\,|\,\alpha_s,\sigma^2) = (2\pi\sigma^2)^{-K/2}\exp(-\frac{1}{2\sigma^2}\|\,y-\Phi\alpha_s\,\|^2) \qquad (5)$$

Where K is the dimension of the observation vector, the compressed sensing problem is transformed into a Bayesian estimation problem. Assuming that Φ is known, the estimated values are sparse vector and noise variance [5; 6].

A priori hypothesis is very important for Bayesian compressed sensing algorithm. Different priori hypothesis describes the prior information of the estimator to be estimated. Making full use of the priori information of the estimator can greatly improve the performance of Bayesian compressed sensing [7; 8].

Speaker recognition algorithm based on bayesian compressed sensing

Speaker recognition based on Bayesian compression perception is to solve sparse representation problem by Bayesian reconstruction algorithm [9-10]. Compared with the existing reconstruction algorithm, bayesian compressive sensing has been proved that, the sparse solution can be obtained even with the interference of observation noise when the real situation is sparse. In the speaker recognition algorithm based on bayesian Mixture perception, the sparse basis is made up of the mean supervector of the Gaussian Mixture Model (GMM).The common Gauss random matrix is selected for the observation matrix and solved by Bayesian compressed sensing algorithm.

In the formula y is the observation value, Φ is the product of sparse basis and Gaussian random matrix, ||α||21 <ζ is the restriction condition of sparse prior. Sparse representation is adopted for speaker recognition, as follows:

Step 1. Training speech and test speech is pretreated.

Step 2. Feature parameters of preprocessed speech frames are extracted.

Step 3. The GMM supervector was calculated, and the over-complete dictionary A was formed by the training speech.

Step 4. solves the sparse decomposition of test speech y in over-complete dictionary a by Bayesian compressed sensing, i.e., y = CA.

Step 5. the residual of the sparse coefficient obtained was calculated by $r_i = \|\,y - \mathrm{A}\,\overset{\wedge}{\delta_i}(x)\,\|_2$.

Step 6. Output discrimination $Identity(y) = \mathrm{argmin}_i(r_i)$.

The GMM supervector is calculated as shown in figure 1.

Experimental results and analysis

In the experiment, the sampling frequency and accuracy are 16 kHz and 16 bit respectively. The recording is carried out in the quiet laboratory environment, and simulated by the MATLAB. The corpus consists of 35 people’s voice, numbered 1 to 35. Fifty voice samples of each person are selected for training, which is about 2 s to 3 s in length. And five samples are used for testing, which is about 10 s in length. 25-dimensional characteristic parameters is formed by the 12-order Mel Frequency Censtrum Coefficient (MFCC), the first-order difference MFCC and pitch period T. The order of GMM mean supervector is set to 64.GMM mean supervectors which constitutes sparse basis, is obtained from the training set data. A test speech supervector is randomly selected.

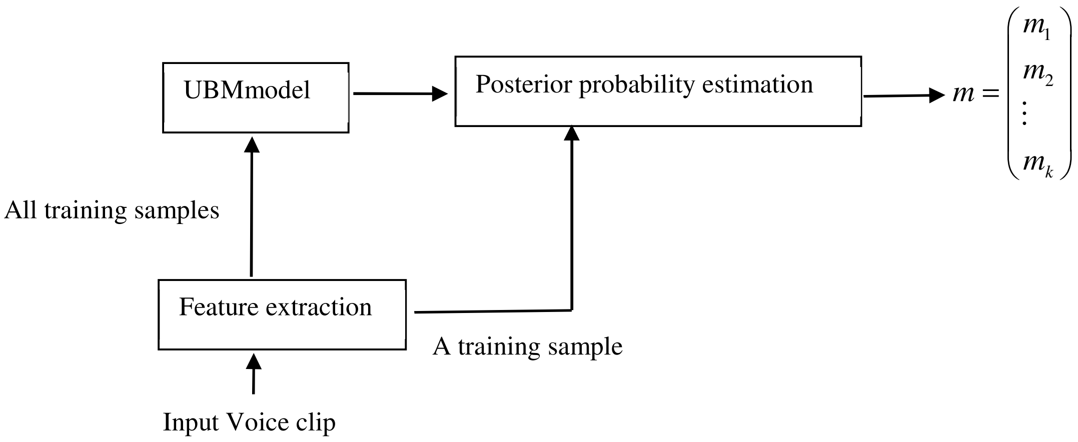
UBMmodel
Posterior probability estimation
$m = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_k \end{pmatrix}$
All training samples
Feature extraction
A training sample
Input Voice clip

Figure 1 – The extraction process of GMM mean supervector

Then after projection by Gauss random matrix, the observed vectors are reconstructed by BCS, and the sparse coefficients are obtained. Then the residual values are calculated. The results are shown in Figure 2.

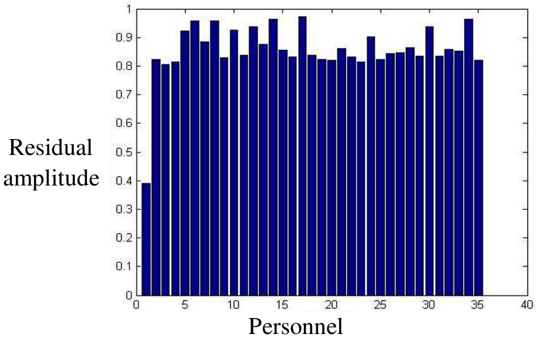

Figure 2 – Residual of speaker recognition algorithm based on BCS

Figure 2 shows the residual of the test hypervector of each speaker in the relative sparse. It is easy to judge from the graph that the test speech belongs to the speaker with number 1. This shows that BCS method can effectively identify the speaker.

The performance of the speakerrecognition system based on Bayesian compression sensing in noise environment is verified and compared with the experimental results of other methods. The signal-to-noise ratio (SNR) of the experiment is 20 dB, 15 dB, 10 dB, 5 dB, 0 dB, -5 dB. In the comparison experiment, the speaker recognition algorithm based on GMM model is selected. The recognition results are shown in Table 1.

From Table 1, it can be seen that in noisy environment the method of CS is better than that of GMM for speaker recognition, and BCS is better than L-1 norm for sparse decomposition. Therefore, compared with CS method, BCS algorithm using multi-level structure prior has a further improvement in recognition rate, especially in the low signal-to-noise ratio, because it uses the Gauss prior whose distribution is closer to the real situation. Compared with the traditional algorithm, there is still a great improvement.

Conclusion

In this paper, Bayesian compression sensing algorithm is used for text-independent speaker recognition, multi-level prior hypothesis is introduced, and sparse solution is solved by using correlation vector machine algorithm in sparse Bayesian learning. Comparing experiments show that Bayesian compressed sensing can obtain a solution closer to 1-norm than Basic Pursuit (BP) algorithm, and has strong robustness.It is proved that BCS is effective in speaker recognition, and also shows the potential of Bayesian compressed sensing in pattern recognition.

Table 1 – The performance of different recognition algorithms in noisy Environment

Recognition methods	20dB	15dB	10dB	5dB	0dB	-5dB
GMM	85.17 %	80.15 %	73.52 %	67.31 %	44.12 %	30.16 %
CS	90.41 %	86.69 %	80.41 %	75.10 %	47.01 %	34.07 %
BCS	91.71 %	86.92 %	81.15 %	76.01 %	48.68 %	35.21 %
ABCS	93.11 %	88.01 %	83.13 %	77.52 %	50.13 %	36.25 %

References

1. *Candes, E, Romberg, J, Tao, T. (2006). Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information [J]. IEEE Transactions on Information Theory, 52(2), 489-509.*
2. *Ji, S., Xue, Y. and Carin, L. (2008). Bayesian Compressive Sensing [J]. IEEE Transactions on Signal Processing, 56(6), 2346–2356.*
3. *Tipping, M.E. (2001). Sparse Bayesian learning and the relevance vector machine [J]. Journal of Machine Learning Research, 1:211–244.*
4. *Jing, Wan, Zhilin, Zhang, Jingwen, Yan, et al. (2012). Sparse Bayesian multi-task learning for predicting cognitive outcomes from neuroimaging measures in Alzheimer's desease [C]. Computer Vision and Pattern Recognition (CVPR), IEEE Conference on Providence, RI, 940–947.*

5. Tipping, M.E. (2001). Sparse Bayesian learning and the relevance vector machine [J]. Journal of Machine Learning Research, 1:211–244.

6. Hong, Sun, Zhilin, Zhang, Lei, Yu. (2012). From sparse to structured sparse: bayesian method [J]. Signal processing, 28 (6), 759-770.

7. Donoho, D.L., Elad, M. (2003). Optimally sparse representation in general dictionaries via l1 minimization [J]. Proceedings of the National Academy of Sciences of the USA, National Acad Sciences, 100(5), 2197.

8. Wipf, D.P. & Rao, B.D. (2004). Sparse Bayesian learning for basis selection [J]. Signal Processing, IEEE Transactions, 52(8), 2153-2164.

9. Wipf, D. & Nagarajan, S. (2010). Iterative reweighted l1 and l2 methods for finding sparse solutions [J]. IEEE Journal of Selected Topics in Signal Processing, 4(2), 317-329.

10. Candes, E.J., Wakin, M.B., & Boyd, S.P. (2008). Enhancing sparisity by reweighted l1 minimization [J]. J Fourier Anal Appl, 14, 877-905.

Стаття надійшла до редколегії 07.03.2019

Ян Ян
Магістр технічних наук, науковий напрямок: електронні комунікації
Яньченський промисловий професійно-технічний коледж школи автомобільної техніки, Яньчен Цзянсу 224005, Китай
Лі Чжао
Школа інформатики та техніки Південно-Східного університету
Нанкін 210096, Китай

ВИВЧЕННЯ АЛГОРИТМУ ПРИЗНАЧЕННЯ СПІКЕРІВ НА ОСНОВІ ТЕОРІЇ ЗБУДЖЕНОЇ БАЙЄСІВСЬКОЇ СИСТЕМИ

Анотація. *Технологія розпізнавання гучномовців широко використовується в Інтернеті та комунікації. Упродовж останніх років теорія стисненого зондування привернула велику увагу в Росії та за кордоном. Порушивши обмеження частоти дискретизації Найквіста, можна вибирати стиснені сигнали, стискаючи їх одночасно з метою підвищення продуктивності системи розпізнавання динаміків. Прагнучи до текстово-незалежної технології розпізнавання динаміків, представлена стаття розкриває поглиблене вивчення алгоритму байєсівського стисненого зондування. Відповідно до характеристик розрідженого коефіцієнта в алгоритмі введено гаусовське попереднє припущення, після чого запропоновано алгоритм розпізнавання динаміків, заснований на байєсівському стислому зондуванні.*

Ключові слова: спікер; визнання; стиснене зондування; рідкісне представлення; BCS GMM

Link to publication

APA *Yan, Yang & Li, Zhao. (2019). A Study on the Speaker Recognition Algorithm Based on Bayesian Compressed Sensing Theory. Management of Development of Complex Systems, 38, 29 – 32, dx.doi.org\10.6084/m9.figshare.9788426.*

ДСТУ *Ян Ян. Вивчення алгоритму призначення спікерів на основі теорії збудженої байєсівської системи [Текст] / Ян Ян, Лі Чжао // Управління розвитком складних систем. – № 38. – 2019. – С. 29 – 32, dx.doi.org\10.6084/m9.figshare.9788426.*