第39卷第10期 中国科学技术大学学报 Vol. 39, No. 10
2009年10月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA Oct. 2009

Article ID: 0253-2778(2009)10-1045-07

Distributed bayesian compressed spectrum sensing based on mutual information

WANG Zhen-xing, YANG Tao, HU Bo

(*Department of Electronic Engineering, Fudan University, Shanghai* 200433, *China*)

Abstract: When compressive sensing is applied in cognitive radio network, spectrum sensing precision by every cognitive radio user is greatly different due to different channel environments between them. Consequently information fusion methods in network and the efficient data processing manner by compressive sensing can be combined to improve sensing precision. First, CS (compressive sampling) is performed independently by every cognitive radio user for rough sensing, and then the sensing information between different users is exchanged for their spatial diversity. Here, mutual information is taken as a measure to evaluate the sensing difference between two cognitive radio users, and those users with large difference are related. The sensing information of every cognitive radio user will be shared under this kind of relationship. After sensing information is shared, Bayesian inference for CS construction in every cognitive radio user is re-built to update the local sensing. The simulation results show that the proposed scheme has advantage both in sensing accuracy and in sensing speed over the conventional scheme.

Key words: cognitive radio; compressed sampling; spectrum sensing; mutual information; bayesian inference

CLC number: TN911.23 **Document code**: A

基于互信息的分布式贝叶斯压缩感知

汪振兴，杨　涛，胡　波

(复旦大学电子工程系，上海 200433)

摘要：压缩感知理论应用于分布式认知网络中时，由于每个认知用户所处的信道环境差别很大，因此频谱感知的精度相差很大．为了提高感知的精度，提出了一种结合了分布式网络中的数据融合方法和压缩感知理论中的高效的数据处理方式的方法．首先，单个认知用户单独地运用压缩采样(CS)进行频谱的粗略感知，然后通过互信息的计算可以得到两两认知用户之间感知信息的差异，而差异大的两个认知用户之间会产生关联．认知用户的感知信息会按照这种关联进行共享．信息共享后，在每个认知用户端，基于贝叶斯推理的压缩感知恢复会重新进行来更新之前的感知结果．仿真结果表明，在感知精度与感知速率方面，算法性能均有改善．

关键词：认知无线电；压缩采样(压缩感知)；频谱感知；互信息；贝叶斯推导

Received: 2009-05-11; **Revised**: 2009-08-28

Foundation item: Supported by National Natural Science Foundation of China (60972024) and the Open Fund of the ASIC & System State Key Laboratory (KF20080402).

Biography: WANG Zhen-xing, born in 1985, master. Research field: cognitive radio. E-mail: 072021046@fudan.edu.cn

Corresponding author: YANG Tao, PhD/associate professor. E-mail: taoyang@fudan.edu.cn

0 Introduction

If the signal is sparse in a certain transform domain, then perfect reconstruction can be acquired through a sub-Nyquist sampling rate, which underlays the currently most innovative signal processing technology — compressive sampling (CS). Since the introduction of CS by D. Donoho[1] and E. Candès, T. Tao[2], a great number of exciting results both in theoretical analysis and experiment have been published. On the theoretical side, Dror Baron[3] proposed the idea of distributed compressive sampling and Lawrence Carin[4] extended generic Bayesian probabilistic method to model compressive sampling. Is there any possibility to fuse these two attractive methods in a distributed cognitive radio system together? This is the intuitive idea behind this paper.

According to compressive sampling theory[1], let $\boldsymbol{X}$ be an $N\times 1$ sparse signal, that is, most of its coefficients are zero, then consider the following measurement system

$$\boldsymbol{Y}=\boldsymbol{HX}+\boldsymbol{n} \tag{1}$$

where $M\times 1$ linear vector $\boldsymbol{Y}$ corresponds to the measurement of the unknown signal $\boldsymbol{X}$, the $M\times N$ measurement matrix $\boldsymbol{H}=[h_1, h_2, \cdots, h_N]$, with $M\ll N$ and $\boldsymbol{n}_{M\times 1}$ represents the acquisition noise as compressive sampling devices do not have infinite precision[2]. Under these conditions, an approximation to the original signal $\boldsymbol{X}$ can be obtained with overwhelming probability by solving the following optimization problem

$$\hat{\boldsymbol{X}}=\arg\min_{\boldsymbol{X}}\{\|\boldsymbol{Y}-\boldsymbol{HX}\|_2^2+\tau\|\boldsymbol{X}\|_1\} \tag{2}$$

where $\|\cdot\|_1$ denotes the $\boldsymbol{l}_1$-norm. Till now a number of methods have been proposed to solve the above CS reconstruction problems, typically including the linear programming algorithm[5,6] and constructive algorithm[7~9].

Cognitive radio (CR) emerged from the fact that the utilization of the current spectrum bands is inefficient. A large span of spectrum for the authorized users stands idle, so these spectrums can be sensed and utilized by cognitive radio sneakily, without any interference to the authorized users. Due to the sparsity existent inherently in these spectrums, CS can be introduced to model the sensing process. Due to the random distribution of the CR users, CS measurements differ greatly from each other. Attenuation, multipath effect etc. can cause measurement errors. This paper proposes a uniform model to characterize these elements and utilize mutual information to evaluate these effects on each cognitive radio user.

1 System model

Consider the following acquisition system in the cognitive radio system

$$\boldsymbol{Y}=\boldsymbol{H}(\boldsymbol{TX}+\boldsymbol{v})+\boldsymbol{n} \tag{3}$$

where $\boldsymbol{X}=[x_1, x_2, \cdots, x_N]^{\mathrm{T}}$ is N-length signal vector in the frequency domain that represents the frequency state in one cell. $\boldsymbol{X}$ has the same value for all cognitive radio users. $\boldsymbol{X}$ depends on how many channels are occupied by the authorized users at one time. $\boldsymbol{T}$ is the transmission matrix that represents characteristic of channel such as attenuation, multipath effect etc. due to different locations of cognitive radio users. $\boldsymbol{H}$ is the compressive sampling matrix statistically generated by cognitive radio user terminal. $\boldsymbol{H}_{M\times N}$ should satisfy the RIP[10] according to compressive sampling theory. This requires that column vectors taken from arbitrary subsets be nearly orthogonal. $\boldsymbol{n}$ is a small amount of noise as compressive sampling devices don't have infinite precision. $\boldsymbol{v}$ is the additive Gaussian channel noise.

Assume $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{n}$ are vector Gaussian signals. Let $\boldsymbol{X}\sim N(0, \boldsymbol{S}_{xx})$, and without loss of generality, assume $\boldsymbol{n}\sim N(0, \boldsymbol{S}_{nn})$. $\boldsymbol{S}_{xx}$ and $\boldsymbol{S}_{nn}$ are covariance matrices of vector Gaussian signal and vector Gaussian noise respectively.

2 Bayesian scheme analysis

In Bayesian modeling, all unknowns are

treated as stochastic quantities with assigned probability distributions.

Expression (3) can be rewritten as

$$\boldsymbol{Y} = \boldsymbol{HTX} + \boldsymbol{\omega} \tag{4}$$

where $\boldsymbol{\omega} = \boldsymbol{Hv} + \boldsymbol{n} = [\omega_1, \cdots, \omega_M]^{\mathrm{T}}$ is a set of independent zero-mean Gaussian variables with variance σ^2. Define $\beta = 1/\sigma^2$, then the likelihood function is formed as

$$p(\boldsymbol{Y} \mid \boldsymbol{X}, \beta) = N(\boldsymbol{Y} \mid \boldsymbol{HTX}, \beta^{-1}) = \left(\frac{2\pi}{\beta}\right)^{-\frac{M}{2}} \exp\left(-\frac{\beta}{2} \| \boldsymbol{Y} - \boldsymbol{HTX} \|_2^2\right) \tag{5}$$

With a Mamma prior placed on β as follows

$$p(\beta \mid a, b) = \Gamma(\beta \mid a, b) \tag{6}$$

The signal $\boldsymbol{X}$ being observed can be assigned a prior distribution

$$p(\boldsymbol{X} \mid \gamma) = \prod_{i=1}^{N} N(x_i \mid 0, \gamma_i), \tag{7}$$

This models our knowledge on the nature of $\boldsymbol{X}$.

Based on the work in Ref. [11], Laplace priors are preferred. In Ref. [12], a new hierachical way was proposed to model the priors by using the following hyper-priors on γ_i,

$$p(\gamma_i \mid \lambda) = \Gamma(\gamma_i \mid 1, \lambda/2) = \frac{\lambda}{2}\exp\left(\frac{-\lambda\gamma_i}{2}\right), \quad \gamma_i \geqslant 0,\ \lambda \geqslant 0 \tag{8}$$

$$p(\boldsymbol{X} \mid \lambda) = \int p(\boldsymbol{X} \mid \gamma) p(\gamma \mid \lambda) \mathrm{d}\gamma = \prod_i \int p(x_i \mid \gamma_i) p(\gamma_i \mid \lambda) \mathrm{d}\gamma_i = \frac{\lambda^{N/2}}{2^N} \exp(-\sqrt{\lambda} \sum_i \mid x_i \mid)$$

Finally take mode λ as the realization of the following Gamma hyper-prior

$$p(\lambda \mid \boldsymbol{v}) = \Gamma(\lambda \mid \boldsymbol{v}/2, \boldsymbol{v}/2) \tag{9}$$

Integrate the above expressions, the joint distribution can finally be defined as

$$p(\boldsymbol{X}, \boldsymbol{Y}, \beta, \gamma, \lambda) = p(\boldsymbol{Y} \mid \boldsymbol{X}, \beta) \cdot p(\beta) \cdot p(\boldsymbol{X} \mid \gamma) \cdot p(\gamma \mid \lambda) \cdot p(\lambda) \tag{10}$$

Where $p(\boldsymbol{Y} \mid \boldsymbol{X}, \beta)$, $p(\beta)$, $p(\boldsymbol{X} \mid \lambda)$, $p(\gamma \mid \lambda)$ and $p(\lambda)$ are defined in Eqs. (5), (6), (7), (8), (9) respectively.

Through iterative updates of hyperparameters, Bayesian inference is performed to obtain estimation of $\boldsymbol{X}$ by MAP standard[13].

3 Mutual information modeling

For a cognitive radio users' network, the aforementioned compressive sampling model can be extended to all nodes. Assume that K cognitive radio users are considered.

$$\left.\begin{aligned} \boldsymbol{Y}_1 &= \boldsymbol{H}_1 \boldsymbol{T}_1 \boldsymbol{X} + \boldsymbol{\omega}_1 \\ &\vdots \\ \boldsymbol{Y}_K &= \boldsymbol{H}_K \boldsymbol{T}_K \boldsymbol{X} + \boldsymbol{\omega}_k \end{aligned}\right\} \tag{11}$$

Assume $\boldsymbol{X}, \boldsymbol{Y}_i, \boldsymbol{n}_i$ are all Gaussian signal vectors and every $\boldsymbol{H}_i$ is an $M \times N$ vector, where $i = 1, \cdots, K$.

$$p(\boldsymbol{X}) \sim \int p(\boldsymbol{X} \mid \lambda) p(\lambda) \mathrm{d}\lambda \tag{12}$$

$$\boldsymbol{\omega}_i \sim N(0, \sigma^2 \boldsymbol{I}) \quad i = 1, 2, \cdots, K$$

Based on the above assumption, mutual information between $\boldsymbol{X}$ and $\boldsymbol{Y}_i$ can be calculated as follows[14]:

$$I(\boldsymbol{X}; \boldsymbol{Y}_i) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(\boldsymbol{X}, \boldsymbol{Y}_i) \ln \frac{p(\boldsymbol{X}, \boldsymbol{Y}_i)}{p(\boldsymbol{X}) p(\boldsymbol{Y}_i)} \mathrm{d}\boldsymbol{X}\mathrm{d}\boldsymbol{Y}_i \tag{13}$$

Where

$$\left.\begin{aligned} p(\boldsymbol{X}, \boldsymbol{Y}_i) &= \iiint p(\boldsymbol{X}, \boldsymbol{Y}_i, \beta, \gamma, \lambda) \mathrm{d}\beta \mathrm{d}\gamma \mathrm{d}\lambda \\ p(\boldsymbol{Y}_i) &= \iiiint p(\boldsymbol{X}, \boldsymbol{Y}_i, \beta, \gamma, \lambda) \mathrm{d}\boldsymbol{X} \mathrm{d}\beta \mathrm{d}\gamma \mathrm{d}\lambda \end{aligned}\right\} \tag{14}$$

and $p(\boldsymbol{X}, \boldsymbol{Y}_i, \beta, \gamma, \lambda)$ is described as Eq. (10).

After each cognitive radio holds its mutual information evaluation value, a matrix can be generated by the following rules.

For every element $m_{i,j}$ in matrix $\boldsymbol{M}_{K \times K}$, its value satisfies:

$$\boldsymbol{M}(i, j) = \mid \boldsymbol{I}(\boldsymbol{X}; \boldsymbol{Y}_i) - \boldsymbol{I}(\boldsymbol{X}; \boldsymbol{Y}_j) \mid, \ 0 \leqslant i, j \leqslant K \tag{15}$$

Apparently, $\boldsymbol{M}$ is a symmetrical matrix. That is, $\boldsymbol{M}^{\mathrm{T}} = \boldsymbol{M}$, where $\boldsymbol{M}^{\mathrm{T}}$ is the transpose form of $\boldsymbol{M}$. Meanwhile,

$$\boldsymbol{M}(i, j) = 0 \quad \text{when } i = j \tag{16}$$

Covert the above matrix to another matrix $\boldsymbol{A}$ through the following way:

$$\boldsymbol{A}(i, j) = \begin{cases} 0, & \boldsymbol{M}(i, j) < \sigma \\ 1, & \boldsymbol{M}(i, j) > \sigma \end{cases} \tag{17}$$

Matrix $\boldsymbol{A}$ can be seen as the connection matrix defined in mathematical graph theory. Here, when $\boldsymbol{A}(i, j)$ equals 1, it means that the ith and the jth cognitive radio users are logical "neighbors". By logical neighbors it is not meant that these two cognitive radio users are located close to each other but that these two cognitive radio users need to exchange individual information. Finally, the network structure is formed (Fig. 1).

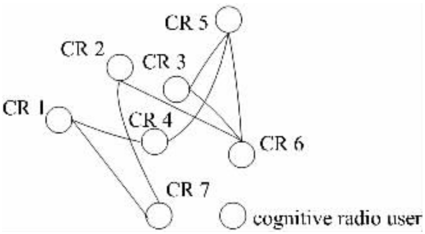

Fig. 1 Illustration of cognitive radio user network

The connection between two cognitive radio users shows the need to exchange individual information between them. So whether two cognitive radio users are linked in the network does not depend on their distance directly, it depends on how many authorized users are located around each cognitive radio user that should be monitored by the cognitive radio user and how far away between the authorized users are from the cognitive radio user.

4 Distributed bayesian inference

In section 3, a graph that represents the connection relationship between cognitive radio users is given (See Fig. 1 for example). This graph determines how these cognitive radio users share their individual information among each other. Any cognitive radio user will transmit its information to those connected with it.

When a cognitive radio user receives information from its neighbors, Bayesian inference is updated through re-computing expressions (5), (6) iteratively.

For illustration, label the collection of "neighbors" of the ith cognitive radio user as $C=\{i_1,\cdots,i_{S_i}\}$, where $i_1,\cdots,i_{S_i}$ are tokens of the corresponding cognitive radio users. S_i is the number of neighbors of the ith cognitive radio user. Then in the ith cognitive radio user, the following equations are available.

$$\left.\begin{aligned} \boldsymbol{Y}_i &= \boldsymbol{H}_i\boldsymbol{T}_i\boldsymbol{X}+\boldsymbol{\omega}_i \\ \boldsymbol{Y}_{i_1} &= \boldsymbol{H}_{i_1}\boldsymbol{T}_{i_1}\boldsymbol{X}+\boldsymbol{\omega}_{i_1} \\ &\vdots \\ \boldsymbol{Y}_{i_{S_i}} &= \boldsymbol{H}_{i_{S_i}}\boldsymbol{T}_{i_{S_i}}\boldsymbol{X}+\boldsymbol{\omega}_{i_{S_i}} \end{aligned}\right\} \tag{18}$$

Integrate these equations to the following one:

$$\left.\begin{aligned} \bar{\boldsymbol{Y}}_i &= \boldsymbol{A}\boldsymbol{X}+\bar{\boldsymbol{\omega}}_i \\ \bar{\boldsymbol{Y}}_i &= \frac{1}{S_i+1}\Big(\boldsymbol{Y}_i+\sum_{j=1}^{S_i}\boldsymbol{Y}_{i_j}\Big) \\ \boldsymbol{A} &= \frac{1}{S_i+1}\Big(\boldsymbol{H}_i\boldsymbol{T}_i+\sum_{j=1}^{S_i}\boldsymbol{H}_{i_j}\boldsymbol{T}_{i_j}\Big) \\ \bar{\boldsymbol{\omega}}_i &= \frac{1}{S_i+1}\Big(\boldsymbol{\omega}_i+\sum_{j=1}^{S_i}\boldsymbol{\omega}_{i_j}\Big) \end{aligned}\right\} \tag{19}$$

Intuitively,

$$\boldsymbol{I}(\bar{\boldsymbol{Y}}_i;\boldsymbol{X}) \geqslant \boldsymbol{I}(\boldsymbol{Y}_i;\boldsymbol{X}) \tag{20}$$

After information shared, Eqs. (5), (6) are updated as follows:

$$p(\bar{\boldsymbol{Y}}_i \mid \boldsymbol{X},\beta_i,\beta_{i_1},\beta_{i_2},\cdots,\beta_{i_{S_i}} = N(\bar{\boldsymbol{Y}}_i \mid \mu_i,\bar{\beta}_i^{-1}) = \left(\frac{2\pi}{\bar{\beta}_i}\right)^{-\frac{M}{2}}\exp\Big(-\frac{\bar{\beta}_i}{2}\|\bar{\boldsymbol{Y}}_i-\mu_i\|_2^2\Big) \tag{21}$$

Where

$$\left.\begin{aligned} \mu_i &= \frac{1}{S_i+1}\Big(\boldsymbol{H}_i\boldsymbol{T}_i+\sum_{j=1}^{S_i}\boldsymbol{H}_{i_j}\boldsymbol{T}_{i_j}\Big)\boldsymbol{X} \\ \bar{\beta}^{-1} &= \frac{1}{(S_i+1)^2}\Big(\beta_i^{-1}+\sum_{j=1}^{S_i}\beta_{i_j}^{-1}\Big) \end{aligned}\right\} \tag{22}$$

Where $p(\beta_i \mid a_i,b_i)\sim\Gamma(a_i,b_i)$, $p(\beta_{i_j} \mid a_{i_j},b_{i_j})\sim\Gamma(a_{i_j},b_{i_j})$, $j=1,2,\cdots,S_i$.

Then $p(\boldsymbol{X},\bar{\boldsymbol{Y}}_i,\bar{\beta}_i,\gamma,\lambda)$ is re-computed through expression (10) using Eqs. (13), (17), (7), (8), (9) instead. After $p(\boldsymbol{X},\bar{\boldsymbol{Y}}_i,\bar{\beta}_i,\gamma,\lambda)$ is calculated, Bayesian inference still should be performed as Ref. [13] describes.

The algorithm procedure is summarized as follows:

Step 1 Each cognitive radio user performs Bayesian compressive sampling independently initially.

Step 2 Through each user's sensing, $p(\boldsymbol{X},\boldsymbol{Y},\beta,\gamma,\lambda)$ is calculated through (10) and a virtual graph is formed as Eq. (17) describes.

Step 3 Transmit information among cognitive radio users following the direction of the graph.

Step 4 After shared information, $p(\boldsymbol{X},\bar{\boldsymbol{Y}}_i,\bar{\beta}_i,$

γ,λ) is re-computed using Eqs. (21), (22), (7), (8), (9), and then Bayesian inference is performed as Ref. [13] describes.

Step 5　Repeat Step 2, 3, 4 until convergence, which means that matrix $\boldsymbol{A}$ in Eq. (17) changes slowly enough.

5　Simulation results and analysis

In this section, we test the performance of the proposed algorithm in the distributed network, in which each user suffers different channel attenuation and distortion.

Simulation parameters are set as follows:

Number of cognitive radio users: 10

Signal length: 512

Number of spikes (sparsity): 80

Firstly, the reconstruction performance is compared between the Bayesian scheme and the Bases Pursuit (BP) scheme, and the simulation is given in Fig. 2. In this scenario, no information between cognitive radio users is shared. The result indicates that generally the reconstruction error for the Bayesian scheme gets smaller as the samples increase, but occasionally a large fluctuation occurs in some samples. Nevertheless, it has an average lower reconstruction error over the BP scheme.

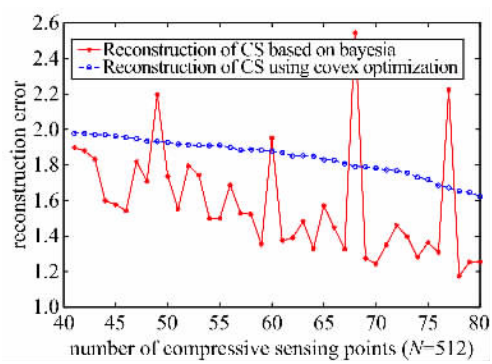

Fig. 2　BP and Bayesian scheme

Next, one cognitive radio user (SNR=10) is selected randomly to verify the effect of cooperation based on mutual information. The improvement on reconstruction performance is given in Fig. 3. Here, "no cooperation" means each cognitive radio user perform compressive sampling as well as the reconstruction independently, "with mutual information" denotes the performance of proposed algorithm, that is, to get information from 4 certain cognitive radio users correlated by mutual information, and "no mutual information" means that the underlying cognitive radio user randomly selects 4 cognitive radio users for cooperation. From Fig. 3, we known that the cooperation scheme based on mutual information greatly eliminates the fluctuation of reconstruction error existing in the Bayesian scheme. Meanwhile, the mutual information scheme can effectively improve the precision of recovery when compared with the randomly selected neighbor scheme. Also, the simulation examines whether more neighbors can bring in better performance. Obviously, in the scenario where the channel quality worsens (SNR=7), a marginal performance improvement can be gained from this kind of cooperation (Fig. 4).

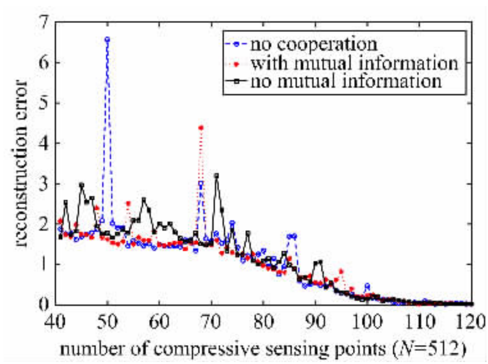

Fig. 3　Effect of cooperation using MI (SNR=10)

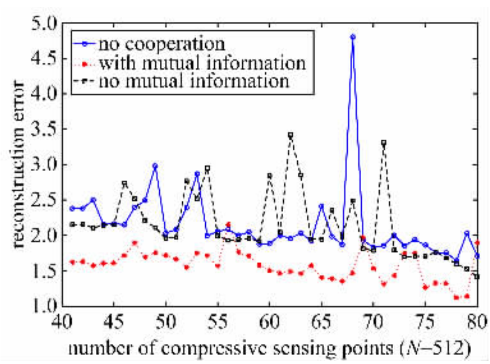

Fig. 4　Effect of cooperation using MI (SNR=7)

In Fig. 5, simulation also examines whether sharing information from more cognitive radio users can bring in better performance. "Reconstruction based on full cooperation" means every cognitive radio user receives information from the other 9 cognitive radio users. "Reconstruction based on MI" means every cognitive radio user only gets information from its neighbors (4 cognitive radio users) via mutual information. From Fig. 4, involving more neighbors for more information does not necessarily offer better precision of reconstruction. Conversely, getting information from unfavorable cognitive radio users makes reconstruction error even more severe.

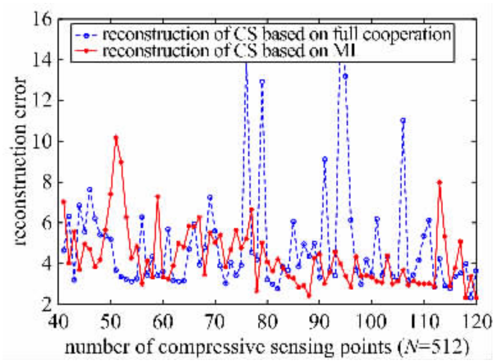

Fig. 5 Full-cooperation and mutual information

Finally, the computation efficiency for the proposed algorithm is analyzed. In practice, the reconstruction of compressive sensing through BP scheme is performed by recursion algorithm until the given recovery precision is reached. BP scheme takes longer in searching for an appropriate base as sampling points increase, whereas, the Bayesian scheme gets the result through an iterative computation framework. We can give a comparison through the simulation in Fig. 6, where the computation time for Bayesian scheme is greatly less than BP scheme and changes slowly as sampling size increases.

The total computation time cost involves two parts: One is the computation of the mutual information between cognitive radio users. The other is the computation of formulas (19), (21), (22). The computing time of these two parts grows approximately linearly as the number of cognitive radio users increases (See Tab. 1). The other time expense of this algorithm is exhibited in interactive communication between linked cognitive radio users, and is a minor part compared with the former ones.

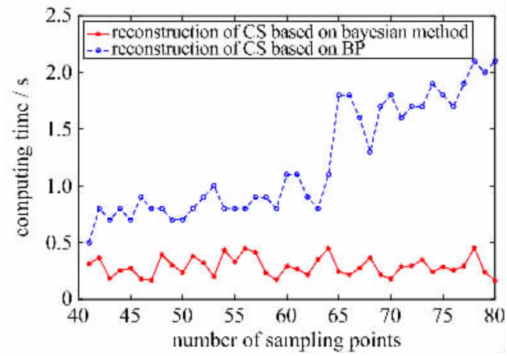

Fig. 6 Computation time for Bayesian and BP scheme

Tab. 1 Algorithm efficiency

number of CR users	time/ms
10	28.5
20	50.5
30	70.0
40	86.7

Besides, since we only consider a slow-varying channel, the procedure described in the end of section 4 converges fast. However, constituting matrix $\boldsymbol{A}$ in Eq. (17), involves computing mutual information using the afore-mentioned formula, which has a relatively high computation complexity.

6 Conclusion

Research on compressive sampling from the Bayesian inference perspective has been proposed in many researchers' works. Here, a novel algorithm is introduced into cognitive radio technology. The new idea that extends Bayesian inference to distributed network is emphasized so that compressive sampling reconstruction is accomplished cooperatively through the shared information among cognitive radio users. Mutual information is preferred for evaluating each cognitive radio user's sensing ability. Experimental results show an excellent performance.

Meanwhile, three major tasks remain to be done: fusing the shared information more effectively, namely, how to generate expression (21), selecting the threshold σ in Eq. (17), and, more importantly, reducing computational complexity.

References

[1] Donoho D L. Compressed sensing [J]. IEEE Transactions on Information Theory, 2006, 52(4): 1 289-1 306.

[2] Candès E J, Tao T. Near optimal signal recovery from random projections: Universal encoding strategies? [J]. IEEE Transactions on Information Theory, 2006, 52(12):5 406-5 425.

[3] Baron D, Duarte M F, Wakin M B, et al. Distributed compressive sensing [DB/OL]. arXiv. org, 2009: arXiv:0901. 3403[2009-06-26]. http://arxiv. org/abs/0901. 3403.

[4] Ji Shihao, Xue Ya, Carin L. Bayesian compressive sensing[J]. IEEE Transactions on Signal Processing, 2008,56(6): 2 346-2 356.

[5] Chen S S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit[J]. SIAM Journal on Scientific Computing, 1998, 20(1):33-61.

[6] Figueiredo M A T, Nowak R D, Wright S J. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems [J]. IEEE Journal of Selected Topics in Signal Processing, 2007,1(4): 586-597.

[7] Tropp J A, Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit [J]. IEEE Transactions on Information Theory, 2007, 53 (12): 4 655-4 666.

[8] Donoho D L, Tanner J. Sparse nonnegative solution of underdetermined linear equations by linear programming [J]. Proceedings of the National Academy of Sciences, 2005, 102(27): 9 446-9 451.

[9] Blumensath T, Davies M E. Gradient pursuits[J]. IEEE Transactions on Signal Processing, 2008, 56(6): 2 370-2 382.

[10] Candès E J, Tao T. Decoding by linear programming [J]. IEEE Transactions on Information Theory, 2005, 51(12): 4 203-4 215.

[11] Seeger M W, Nickisch H. Compressed sensing and bayesian experimental design[C]// Proceedings of the 25th International Conference on Machine Learning. New York: ACM, 2008: 912-919.

[12] Figueiredo M A T. Adaptive sparseness for supervised learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003,25(9):1150-1 159.

[13] Babacan S D, Molina R, Katsaggelos A K. Beyesian compressive sensing using Laplace priors [J/OL]. IEEE Transactions on Imaging Processing, 2009; doi: 10. 1109/TIP. 2009. 2032894 [2009-05-01]. http:// http://ieeexplore. ieee. org/search/freesearchresult. jsp? history = yes&queryText =% 28bayesian + compressive+sensing+using+laplace+priors%29.

[14] Shannon C E. A mathematical theory of communication [J]. Bell System Technical Journal, 1948,27: 379-423.