Applying Compressive Sensing to the Cocktail Party Problem

Eli Saracino
Department of Computer Science
Boston University
Boston, MA 02135
esaracin@bu.edu

Andy Huynh
Department of Computer Science
Boston University
Boston, MA 02135
ndhuynh@bu.edu

Abstract

A Compressed Sensing framework to approach blind source separation is investigated. We compare previous attempts, such as Nonnegative Matrix Factorization, and convey the potential advantages of Compressive Sensing: namely, it providing a more accurate reconstruction that relies far less on the necessity of learning.

1 Introduction

The applications of the Cocktail Party Problem are far-reaching: from surveying and separating radio signals, to imagining and examining neural signals, source separation plays an important role in an ever-expanding number of fields. An efficient approach to solving the Problem, then, has implications that extend far beyond the realm of Computer Science.

Past techniques applied to solve this problem have, at a high level, relied largely on the learning of a specific dictionary, and the utilization of this dictionary to reconstruct distinct signals from potentially mixed sources. One such approach is Nonnegative Matrix Factorization, hereby NMF. As suggested, this method implies that prior information about speech sources is known in order to work properly, and, perhaps even more to its detriment, does not provide a well-defined solution in the case of over-complete dictionaries, which are so often utilized in the Compressive Sensing framework to minimize the number of measurements needed to reconstruct a given signal. (cite 3) The NMF problem statement can be given as follows:

$$E = ||\mathbf{Y} - \bar{\mathbf{D}}\mathbf{H}||_F^2 + \lambda \sum_{ij} \mathbf{H}_{ij} \text{ s.t } \mathbf{D,H} \geq 0$$

Notably, the problem statement is not totally dissimilar to that of the standard Compressive Sensing model, albeit with some minor differences.

History and theory

Using Compressive Sensing, we can model this problem in a new light. To do this, we first have make an assumption of sparsity; that is, in analyzing a mixed audio signal, at any given instant in time, we must assume that noise is coming from only a single source. In this way, the *frequency* domain of a mixed signal will be sparse, even if it should be the case that the the signal is quite robust in the time domain. With this assumption, we can frame the problem can be as a case of sparse signal recovery, and is solvable using any one of several greedy algorithms. Specifically, we use Orthogonal Matching Pursuit to reconstruct the separate sparse signals.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

2 Methods

Methods section (use paper)

3 Results

results (comparison to other techniques (ICA))

References

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Schmidt, Mikkel N. and Rasmus Kongsgaard Olsson. "Single-channel speech separation using sparse non-negative matrix factorization." *INTERSPEECH* (2006).

[4] BAO, G., YE, Z., XU, X. AND ZHOU, Y. *A Compressed Sensing Approach to Blind Separation of Speech Mixture Based on a Two-Layer Sparsity Model* - IEEE Journals & Magazine Bao, G., Ye, Z., Xu, X. and Zhou, Y. (2017). A Compressed Sensing Approach to Blind Separation of Speech Mixture Based on a Two-Layer Sparsity Model - IEEE Journals & Magazine. [online] Ieeexplore.ieee.org. Available at: http://ieeexplore.ieee.org/document/6384713/