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 efficient 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 in a medical setting, 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. **[1]** The NMF problem statement effectively tries to optimize the following cost function:

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

where $\bar{D}$ is a column-wise normalized dictionary matrix and $\mathbf{H}$ is the sparse solution upon which an L_1 norm penalty is induced. In keeping with the theme of NMF, both $\bar{D}$ and $\mathbf{H}$ must be nonnegative. In the above statement, λ is a parameter that controls the degree of sparsity.

Notably, the problem statement is not totally dissimilar to that of the standard Compressive Sensing model, albeit with some minor differences. For example, the dictionary $\bar{D}$ must be learned, usually by training on a large dataset of a single speaker. This, of course, sacrifices efficiency and casts framework of the reconstruction as a learning problem.

With the advent of Compressive Sensing, however, 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 as a case of sparse signal recovery, defined as follows:

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

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$$

where **S** is our set of source signals and **X** is a mixture of these signals as determined by our mixing matrix, **A**. As we will go on to show, determining **A** becomes a key step in the Compressive Sensing approach to solving the problem.

While solving this problem subject to the minimization the L_0 norm of **s** is notably NP-hard, we can obtain an approximate solution using any one of several greedy algorithms. Specifically, we apply Orthogonal Matching Pursuit to reconstruct the separate sparse signals, and transform them back into the time domain to complete the recovery.

In the coming sections, we will describe in more detail the compressive sensing approach used and it's computational benefits over such techniques as NMF, as well as some of its own shortcomings.

2 Methods

The following is an overview of the algorithmic steps we take in applying Compressive Sensing to the Cocktail Party problem to reconstruct distinct signals from mixed input sources.

In keeping with our assumption of a frequency-sparse input signal, the first step we take is to compute the Short-time Fourier transform of our source signal in order to cast it into its time-frequency domain. Shown below is a figure that demonstrates the utility of this transformation: an input signal that was original robust becomes relatively sparse, with distinct structure forming along specific axes.

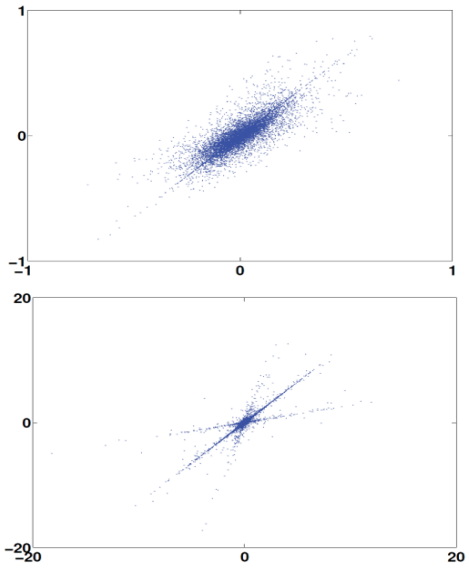

Figure 1: Above, a mixed signal in the Time domain. Below, in the Transform domain

In particular, notice the sp

Methods section (use paper)

3 Results

results (comparison to other techniques (ICA))

References

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

[2] 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/