Blind source separation of mixed audio signals

Venkatesh Bharadwaj Srinivasan

December 23, 2020

Introduction

Imagine a scenario where there are multiple people speaking and you would like to listen to one person at a time to catch hold of what exactly he or she is saying. Our primary goal is to separate individual audio sources from a mixture, where we are not given any details about the sources or the kind of mixing. We make use of a technique named Independent Component Analysis to separate the mixed audio signals into individual components.

Independent component analysis (ICA) [1] is a statistical method for transforming an observed multidimensional random vector into components that are as statistically independent as possible from each other, for revealing the hidden factors that are within the random signals. To implement this, we need to have the number of individual sources to be equal to that of the total number of mixtures. Using superposition principle, each mixture is made to contain a different linear combination of individual sources. For the sake of simplicity, let's discuss only about binary ICA, which makes use of 2 sources. We are basically given two linear mixtures of two source signals which we know to be independent of each other, i.e. observing the characteristics of one signal does not give any information about the other. We had adopted a fixed point algorithm technique called FASTICA which uses an orthogonal rotation of uncorrelated data and maximizes a measure of non-Gaussianness of the rotated components. Non-gaussianness serves as a very strong condition and requires infinite data to verify.

We model this problem using the equation $\mathbf{x} = \mathbf{A}\mathbf{s}$, where $\mathbf{x}$ is the observed mixed signal data, $\mathbf{A}$ is the mixing matrix contains the coefficients of the linear transformation of the source signals and $\mathbf{s}$ is the latent vector that contains the independent source signals. We assume that the independent source signals have non-Gaussian distributions. We center the $\mathbf{x}$ matrix by subtracting its mean from the observed mixed signal data. We make use of Singular Value Decomposition to break the observed matrix into 3 parts. We then compute $\mathbf{U}$ and Σ by estimating the covariance of the centered data. We then force the signals to be uncorrelated via a linear transformation $\hat{s} = Vx_w$, (where $x_w = \Sigma^{-1}U^T x$) through a process called 'whitening'. This step is done by obtaining the unmixing matrix $W = V\Sigma^{-1}U^T$. We estimate the appropriate orthogonal vector V by means of maximum likelihood estimation of the entropy of Vx_w, which maximizes the non-Gaussianness of the individual signals. This is done because, as per central limit theorem, a linear combination of non-Gaussian is more gaussian than the individual signals. Then, we identify the riwght $\hat{s}$ using the estimated V.

This algorithm was tested for a few mixed signal audio files and the separation of independent signals from the binary source mixture was obtained. PFA the poster and results[2].

[1] https://www.cs.helsinki.fi/u/ahyvarin/papers/NN00new.pdf

[2] https://tinyurl.com/soundzoned