

Resumen

El avance tecnológico en los mecanismos de obtención de datos en multitud de materias ha provocado que la base de datos disponible sea muy extensa y contenga una mayor información que podría llevar a la resolución del problema en estudio. Dicha información se encuentra latente a la espera de ser descubierta pero dada la magnitud de la misma, el análisis y la extracción de características se hace cada vez más difícil. La forma en la que los datos se representen parece de vital importancia a la hora de conseguir sacar a la luz estas estructuras ocultas.

La descomposición de las observaciones en un nuevo conjunto de menor dimensión sin necesidad de pérdida de información relevante juega un papel fundamental a la hora de mejorar la extracción de las variables latentes. El *Análisis de Componentes Principales (PCA)*, la *Descomposición en Valores Singulares (SVD)*, la *Factorización No Negativa de Matrices y Tensores (NMF/NTF)* constituyen los mecanismos de descomposición lineal y multilineal de las observaciones que permiten representar la información de una manera más práctica para su utilización.

A lo largo del proyecto realizaremos un estudio de estos métodos de descomposición de las observaciones que nos servirá de base para la implementación del algoritmo *ThinICA* con el que abordaremos la resolución del problema de la *Separación Ciega de Fuentes (BSS)*.

Este algoritmo se pondrá en práctica con señales de diferente naturaleza (audio, electrocardiogramas, modulaciones digitales, imágenes naturales y artificiales) procedentes de algunos de los campos de actuación en los que la *BSS* tiene cabida. La calidad de la extracción alcanzada será medida a través de la comparación de los índices de prestaciones obtenidos con los algoritmos de referencia existentes (*JADE*, *AMUSE* y *SOBI*).

El objetivo final del presente proyecto consiste en completar las tres etapas que plantea la separación ciega de fuentes: un estudio teórico del problema, la implementación de una solución y la comprobación de resultados a través de experimentos prácticos.

Abstract

Technological advances in the mechanisms of datasets collection grow up the databases available. Such datasets have underlying latent components or factors that could help solving the studied problem. But as the datasets grow up the hidden variables are getting harder and harder to extract. The way that the data is represented plays a fundamental role in enhancing the data and extracting hidden components.

To replace the original data by a lower dimensional approximate representation maintaining the meaningful information seems to be a key role improving the extraction of the latent variables. *Principal Components Analysis (PCA)*, *Singular Value Decomposition (SVD)* and *Nonnegative Matrix and Tensor Factorizations (NMF/NTF)* are linear and multi-linear decomposition models which enhance the representation of data.

Throughout this project we will study the decomposition methods of the observations that will help us in the implementation of the *ThinICA* algorithm. This algorithm will be used to work out the solution of the *Blind Source Separation (BSS)* problem.

This algorithm will be played in signals of different nature (audio, electrocardiogram, digital modulations and nature/artificial images) from various of the fields from which the *BSS* problem acts. The measurement of the quality will be made by the comparison of the performance indices studied with the reference algorithms (*JADE*, *AMUSE* and *SOBI*).

The goal of this project is to complete the three-stages problem of the blind source separation: a theoretical study of the problem, an implementation of a solution and a verification of the results obtained in the experiments.

