ECOPICA: Empirical Copula-Based Independent Component Analysis

畢紘愷

Abstract

Independent component analysis (ICA) is an important unsupervised learning statistical method in data analysis and has been widely used in different fields. The main goal of ICA is to find a transformation to recover latent independent variables from observations. Although some well-known algorithms implicitly assume that the source distribution is symmetric, this assumption is sometimes invalid for real data. Some studies have further pointed out that these well-known algorithms cannot properly recover the source signal when the source distribution is skewed and its kurtosis is close to that of the Gaussian distribution.

In this study, we propose a non-parametric ICA method, called ECOPICA, which is based on the empirical copula of data. Unlike some well-known algorithms, this method does not require any assumptions regarding source distribution. We use the distance between the empirical copula and the uniform copula relative to the distance between the maximum copula and the uniform copula to measure the dependency and use the grasshopper optimization algorithm to minimize this dependence. Our simulation and empirical analysis show that compared with other well-known ICA approaches, ECOPICA performs well under the assumption of different source distribution shapes. In particular, when the source distribution is skewed and its kurtosis is close to that of the Gaussian distribution, the performance of ECOPICA is significantly better than that of other well-known models. To improve the implementation efficiency, we adopt a mini-batch trick to substantially reduce the computational complexity and implement the algorithm with a high-performance parallel algorithm of C++. Finally, we propose a gradient-based optimization concept as an outlook for future research.

Keywords: Independent component analysis (ICA), copula, kurtosis, skewness, grasshopper optimization algorithm (GOA), cocktail-party problem, blind image separation.