On Improved Methods for Blind Separation of Signals with Arbitrary Probability Distributions and Robust Independent Component Analysis

著者	Muhammad Tufail
号	51
学位授与番号	3669
URL	http://hdl.handle.net/10097/37337

氏　　　　　名	Muhammad Tufail
授　与　学　位	博士（工学）
学位授与年月日	平成１８年９月１３日
学位授与の根拠法規	学位規則第４条第１項
研究科，専攻の名称	東北大学大学院工学研究科（博士課程）電子工学専攻
学位論文題目	On Improved Methods for Blind Separation of Signals with Arbitrary Probability Distributions and Robust Independent Component Analysis（任意の確率分布を持つ信号のブラインド分離のための改良法とロバスト独立成分分析について）
指　導　教　員	東北大学教授　川又　政征
論文審査委員	主査　東北大学教授　川又　政征　東北大学教授　吉澤　誠 東北大学教授　青木　孝文　東北大学講師　阿部　正英

論文内容要旨

Abstract: **In this thesis, we concentrate on two important issues related to blind source separation (BSS) using independent component analysis (ICA). First, we note that the separation performance of various ICA estimation methods depends on the probability distributions of source signals. It is therefore important to have algorithms that work well for signals with a wide range of pdfs. Second, it is highly preferable in practical applications to employ a method that is robust in the presence of outliers.**

In order to address the first problem, we propose two different approaches. The first approach is based on employing certain parametric nonlinear activation functions in the relative gradient algorithm, which is one of the widely used ICA estimation method. The parameters of these activation functions are adapted online so as to ensure that the desired solution remains a locally stable equilibrium point of the algorithm. Numerous simulation results are presented to demonstrate that the proposed methods give good separation performance for signals with many different probability distributions.

A central principle for performing BSS by ICA is maximization of non-Gaussianity. Non-Gaussianity is generally (quantitatively) measured by employing a non-quadratic function. However, as we mentioned above, the separation performance depends on the probability distributions of sources and consequently a single nonlinear function may not give adequate results for all signals. Accordingly, we propose to employ an empirical characteristic function based non-Gaussianity measure that directly computes the distance of an arbitrary distribution from the Gaussian in the characteristic function domain. The proposed non-Gaussianity measure is a weighted distance between the characteristic function of a random variable and a Gaussian characteristic function at some appropriately chosen sample points. We derive a fixed-point algorithm to optimize this non-Gaussianity measure and also suggest a procedure to choose adequate sample points online so as to improve the separation performance. Computer experiments with many different types of sources show that the characteristic function based non-Gaussianity measure has the potential to give adequate performance for a wide variety of distributions.

Finally, in order to address the problem of robustness, we propose to employ an extension to the natural gradient algorithm in which an exponentially decaying function is introduced to discount the effect of outliers. By appropriately choosing the spread of the exponential function, the separation performance is greatly improved in the presence of outliers; the performance remains almost the same when there are no outliers in the data.

I. INTRODUCTION

Blind source separation (BSS) refers to the problem of recovering a set of unobserved source signals from their observed linear mixtures with unknown mixing coefficients. The observed signals are generally obtained at the outputs of different sensors and are almost always corrupted by additive noise. The mixing process, also called the (noisy) BSS model, can be expressed in a compact form as

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{v}(k), \qquad 1$$

where $\mathbf{x}(k)=[x_1(k),x_2(k),...,x_m(k)]^T$ is an m-dimensional vector of observed signals at time instant k, $\mathbf{s}(k)=[s_1(k),s_2(k),...,s_n(k)]^T$ is an n-dimensional vector of source signals, $\mathbf{A}$ is an $m\times n$ ($m\geq n$) full rank mixing matrix and $\mathbf{v}(k)=[v_1(k),v_2(k),...,v_m(k)]^T$ is a vector of noise signals present at m sensors. The noise vector $\mathbf{v}(k)$ is often assumed to be Gaussian and statistically independent to the individual sources.

The objective of BSS is to determine a separating matrix $\mathbf{W}=\hat{\mathbf{A}}^{\dagger}$ of dimension $n\times m$ such that the n-dimensional transformed vector $\mathbf{y}(k)=\mathbf{W}\mathbf{x}(k)$ given by

$$\mathbf{y}(k)=\mathbf{W}\mathbf{A}\mathbf{s}(k)+\mathbf{W}\mathbf{v}(k)=\mathbf{H}\mathbf{s}(k)+\mathbf{W}\mathbf{v}(k), \quad 2$$

estimates the original input vector $\mathbf{s}(k)$ given only the observed vector $\mathbf{x}(k)$ and certain assumptions about the statistics of source signals. Here $\hat{\mathbf{A}}^{\dagger}$ is an estimate of the (pseudo) inverse of $\mathbf{A}$ and matrix $\mathbf{H}=\mathbf{W}\mathbf{A}$ of dimension $n\times n$ is the global transformation matrix from $\mathbf{s}(k)$ to $\mathbf{y}(k)$.

Over the last decade, BSS has received a considerable research attention due to its large number of application areas including biomedical signal processing, audio and speech signal separation, image processing and features extraction, telecommunications, and financial applications [1], [2]. Consequently, numerous algorithms have been proposed in literature to tackle the problem of BSS. These algorithms can be categorized based on the assumptions they make on the statistics or diversities of the source signals. A particular algorithm works only for those signals that satisfy the assumptions (or separation conditions) of that algorithm. In particular, ICA solves the problem of BSS provided the following conditions are satisfied [1]

- The source signals are mutually statistically independent, and
- At most one source is Gaussian.

Blind separation of signals based (only) on the above two assumptions is equivalent to transforming the observed vector into another vector with (maximally) statistically independent components. These independent components are the estimates of individual sources. The ICA methods perform BSS by (directly or indirectly) reducing dependencies between the reconstructed signals. Since independence is a much stronger condition than simple decorrelation, HOS are employed in order to achieve BSS by ICA.

II. MOTIVATION

The main topic of this thesis is blind separation of instantaneous mixtures (of sources) using ICA. As mentioned earlier, ICA solves the BSS problem by making a strong but physically plausible assumption that the original (non-Gaussian) sources are mutually statistically independent. Based on this assumption, BSS is accomplished by obtaining a de-mixing matrix $\mathbf{W}$ such that the components of the corresponding output vector $\mathbf{y}(k)$ become as independent as possible. It may be noted that unlike simple decorrelation, HOS are indispensable in order to realize statistical independence between random variables. Consequently, various algorithms for ICA employ HOS (or nonlinear activation functions that inherently incorporate HOS) in an attempt to reduce higher-order correlations between the components of the output vector.

The choice of nonlinear activation functions mentioned above is critical since the local stability and the statistical efficiency of different learning algorithms (for ICA) depend on these nonlinear functions (with respect to the probability distributions of sources). The best separation performance is attained if we choose these nonlinearities as the score functions of sources [3]. The score function is defined as

$$\varphi=\frac{p'_{s_i}}{p_{s_i}} \quad 3$$

Where p_{s_i} is the probability density function (pdf) of the source s_i. This implies that in order to use score functions as activation functions, we need to have knowledge about the pdfs of sources. However, such information is rarely available in practical applications. Therefore, in these situations, it is necessary to obtain the score functions / adequate nonlinearities from the observed data or employ an ICA method that works well for signals with a wide range of pdfs. This is precisely the first issue we consider in this thesis.

Another problem we address in this thesis is related to ICA of signals when there are potential outliers in the observed data. An outlier is an observation that lies far away from the rest of the data. Such extreme observations can have an unsettling influence on the learning algorithm (for ICA) especially when the activation functions are fast growing nonlinearities. It is therefore important to identify and either accommodate or reject these outliers so as to reduce their influence on the learning algorithm.

In short, we address the following two issues in this thesis

- BSS / ICA of signals with arbitrary probability distributions, and
- Robust ICA in the presence of outliers.

III. PROPOSED METHOD-I: PARAMETRIC NONLINEAR ACTIVATION FUNCTIONS

In Chapter 3, we propose a simple method for online estimation of activation functions in order to blindly separate instantaneous mixtures of sub-Gaussian and super-Gaussian signals. An adequate choice of these activation functions is necessary not only for a successful source separation (using relative gradient algorithm), but also to achieve sufficient level of cross-talk index. To accomplish this, we employ a simple parameterized model for the probability density functions of sources [4]. The parameter of this distribution model (for each estimated source signal) is adapted online by maximizing the log-likelihood, while the activation functions are obtained as the associated score functions. Furthermore, a modified relative gradient algorithm is derived that exhibits an isotropic convergence (near the desired solution) independent of the statistics of sources. Some simulation results are given to demonstrate the effectiveness of the presented methods.

Another method we propose in Chapter 3 is based on exploiting exponential type nonlinearities in order to blindly separate instantaneous mixtures of source signals [5]. These nonlinear functions are applied only in a certain range around zero in order to ensure the stability of the separating algorithm. The proposed truncated nonlinearities neutralize the effect of outliers while the higher order terms inherently present in the exponential function result in fast convergence especially for signals with bounded support. By appropriately varying the truncation threshold parameters, signals with both symmetric and asymmetric probability distributions can be separated. In case of symmetric distributions, we need to tune only one threshold parameter. An adequate value of this parameter depends on whether the source signal is sub-Gaussian or super-Gaussian. When the sources consist of signals with mixed kurtosis signs, we estimate the value of normalized kurtosis online in order to classify the signals as sub-Gaussian or super-Gaussian and consequently choose an adequate value of the truncation threshold. For asymmetric pdfs, we can employ a procedure similar to the one described above but with two truncation thresholds. The selections of appropriate thresholds (and nonlinearities) are made by estimation the left and the right normalized kurtosis values of the outputs.

IV. PROPOSED METHOD-II: CHARACTERISTIC FUNCTION BASED INDEPENDENT COMPONENT ANALYSIS

A central principle for estimating the ICA model is maximization of non-Gaussianity. Specifically, ICA of an observed random vector can be performed by making the outputs of the de-mixing system as non-Gaussian as possible. In order to implement this approach for ICA estimation, we need some quantitative measure of non-Gaussianity. A natural criterion for non-Gaussianity is negentropy, which attains its minimum (for constant variance) when the distribution is Gaussian and all other distributions have larger negentropies. However, the problem in using negentropy is that it is computationally very difficult and requires an estimate of the pdf. Therefore, in practice, we must use some (simpler) approximations of negentropy.

The classical method of approximating negentropy is based on expanding (like a Taylor expansion) an arbitrary pdf in the vicinity of the Gaussian density. The resultant approximations are given in terms of different combinations of higher-order cumulants including kurtosis. Higher-order cumulants, nevertheless, have some shortcomings, particularly when their values have to be estimated from a measured sample. First, finite-sample estimators of higher-order cumulants are highly sensitive to outliers. Secondly, perfectly estimated cumulants measure mainly the tail of the distributions and are largely unaffected by structure around the center of distributions. These drawbacks of higher-order cumulants can be avoided by approximating negentropy using nonpolynomial functions. Although, practically any smooth non-quadratic function can be employed for this purpose, the optimal nonlinear functions depend on the (unknown) pdfs of sources [6]. This implies that if some preselected (parametric) nonlinearity differs considerably from the optimal function, the associated learning algorithm may perform poorly for these sources.

It is therefore necessary to employ some direct measure of non-Gaussianity that works well for signals with a wide range of pdfs. In Chapter 4, we propose to utilize an empirical characteristic function (ecf) based non-Gaussianity measure (contrast function) in order to perform ICA [7]. It may be noted that ecf being the Fourier-Stieltjes transform of the empirical distribution retains all information about the data. Consequently, the estimation methods based on the ecf can be made as efficient as the likelihood-based approaches. In ICA estimation, the ecf has already been used in [8] to construct an objective function for measuring statistical independence between random variables. In Chapter 4, we employ the ecf to (directly) measure the distance of an arbitrary empirical distribution from the Gaussian distribution (at some adequately chosen sample points). Such a contrast function can be easily maximized by using a fixed-point algorithm. Furthermore, we also suggest a procedure for choosing appropriate sample points (from an initially chosen sample vector) in order to obtain somewhat better separation performance. Finally, some simulation results are given in order to show that the proposed approach works well for both symmetric and asymmetric distributions.

V. PROPOSED METHOD-III: ROBUST INDEPENDENT COMPONENT ANALYSIS

In Chapter 5, we propose an extension to the natural gradient algorithm for robust ICA in the presence of outliers [9]. The standard natural gradient algorithm does not exhibit this robustness property since it is associated to an estimating function that is not a bounded function of the inputs for any choice of nonlinear activation functions. Consequently, the existence of only a few outliers may have an unsettling influence on the algorithm. In the proposed approach, an exponential type decaying function is introduced that gives smaller weights to potential outliers so that their influence on the learning algorithm is weakened. If we employ monomial functions as nonlinear activation functions, the proposed approach becomes equivalent to replacing ordinary sample moments (in the natural gradient algorithm) with robust higher-order moments. Computer simulations are presented to show that the proposed method, as compared to the standard natural gradient algorithm, gives better separation performance in the presence of outlying data; the performance remains almost the same when there are no outliers in the data.

V. CONCLUSIONS

In this thesis, we have proposed some methods in order to obtain an improved separation performance for signals with a wide variety of pdfs and / or when there are potential outliers in the data. Numerical simulations are provided to highlight the potential performance gains of the proposed methods.

REFERENCES

[1]. A. Cichoki, and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, England, John Wiley and Sons, Ltd., 2002.

[2]. A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, New York, John Wiley and Sons, Inc., 2001.

[3]. J. -F. Cardoso, and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017-3030, 1996.

[4]. M. Tufail, M. Abe and M. Kawamata, "Blind separation of statistically independent signals with mixed sub-Gaussian and super-Gaussian probability distributions," *Proceedings of IEEE International Symposium on Circuits and Systems*, Japan, pp. 3027-3030, May, 2005.

[5]. M. Tufail, M. Abe and M. Kawamata, "Blind separation of mixed kurtosis signals using local exponential nonlinearities," *Proceedings of IEEE International Midwest Symposium on Circuits and Systems*, USA, pp. 39-42, August, 2005.

[6]. A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626-634, 1999.

[7]. M. Tufail, M. Abe and M. Kawamata, "A characteristic function based contrast function for blind extraction of statistically independent signals," *IEICE Transactions on Fundamentals of Electronic, Communications and Computer Sciences*, vol. E89-A, no. 8, pp. 2149-2157, 2006.

[8]. J. Eriksson, and K. Koivunen, "Characteristic-function-based independent component analysis," *Signal Processing*, vol. 83, no. 10, pp. 2195-2208, 2003.

[9]. M. Tufail, M. Abe and M. Kawamata, "An extension to the natural gradient algorithm for robust independent component analysis in the presence of outliers," *IEICE Transactions on Fundamentals of Electronic, Communications and Computer Sciences,* (in press).

論文審査結果の要旨

独立成分分析（ICA）はランダムなベクトル信号を統計的に独立な成分をもつ別なベクトル信号に線形変換しようとする信号処理手法であり，その典型的な応用はブラインドソース分離（BSS）である．この論文では ICA を使った BSS における二つの重要な課題を考察している．一つはさまざまな確率分布関数をもつ信号に対して働くアルゴリズムを見出すことであり，他の一つは外れ値がある場合において頑健なアルゴリズムを見出すことである．本論文は，以上の二つの課題に関して考察し，BSS の性能向上について論じたものであり，全編６章よりなる．

第１章は緒言であり，本論文の対象である BSS の目的と課題についてまとめている．

第２章では，本論文において必要とされる数学的概念を紹介し，ICA モデルを推定するための従来法を概観している．特に，勾配に基づく方法について詳しく論じている．

第３章では，ICA のためのパラメトリックな非線形活性化関数を提案している．この非線形関数のパラメータは，学習アルゴリズムが局所的に安定となることを保証するように，オンラインで適応的に変化する．これらのパラメータのための適応アルゴリズムの詳細な記述を与え，特に，提案された打ち切り指数関数は，打ち切りしきい値パラメータを適切に選ぶことによって，対称および非対称確率分布をもつ信号源を分離できることを示している．数値シミュレーションの結果により，提案法は様々な確率分布をもつ信号に対して，従来法に比べて約１／２の繰り返し回数で分離性能指数が収束することを示している．これは実用上，重要な成果である．

第４章では，ICA による BSS を行うための中心的な原理は非ガウス性を最大化することである点を考慮し，非ガウス性測度に基づく経験特性関数を利用することを提案している．この非ガウス性測度を最適化するために不動点アルゴリズムを導出し，オンラインで十分な標本点を選ぶための手順を提案し，その結果，分離性能を向上させている．様々なタイプの信号源を用いた数値シミュレーションにより，非ガウス性測度に基づく特性関数が様々な分布関数の信号に対して分離性能指数を向上させることを示している．

第５章では，信号の分布関数に対する頑健性の問題を扱うために，指数関数的に減少する関数を用いて外れ値の影響を除外し，拡張された自然勾配アルゴリズムを利用することを提案している．指数関数の広がりを適切に選ぶことによって，分離性能指数は，外れ値がある場合には大きく改善され，外れ値がない場合にはほとんど劣化しないことを数値シミュレーションによって示している．これは重要な知見である．

第６章は結言である．

以上要するに本論文は，任意の確率分布を持つ信号の BSS のために，新しい非線形活性化関数，非ガウス性測度，外れ値を除外する指数関数を提案し，BSS の性能の向上を図ったものであり，電子工学および信号処理工学において寄与するところが少なくない．

よって，本論文は博士(工学)の学位論文として合格と認める．