

深層パーミュテーション解決法の基礎的検討

Basic study for deep permutation solver.

蓮池 郁也

Fumiya Hasuike

1 はじめに

音源位置やマイクロフォン位置等が未知の条件で音源分離を達成する技術はブラインド音源分離 (BSS) と呼ばれる。観測信号のチャンネル数 (マイクロフォン数) と混合されている音源数が等しい条件下では、観測信号を時間周波数領域に変換し周波数毎に独立成分分析 (ICA) [1] を適用する時間周波数領域 ICA (FDICA) [2] が提案されている。ここで、ICA は一般に推定分離信号の順番が不定であり、FDICA は周波数毎に独立な ICA による BSS を行うため、分離信号の順番が周波数毎にばらばらになってしまう問題が生じる。この問題は一般に「パーミュテーション問題」と呼ばれている。

本稿では、前述の問題を解決するために深層学習 (DNN) に基づくパーミュテーション解決法を検討する。また、本論文で提案する深層パーミュテーション解決法を人工データと実際の音声及び音楽信号に適用し、DNN に基づくパーミュテーション問題解決の可能性について調査する。

2 従来手法

2.1 FDICA とパーミュテーション問題

FDICA は音源間の統計的独立性のみに基づいて分離行列を推定するため、分離音源のスケール及び順番に関して推定することができない。従って、FDICA の推定分離行列を \hat{W}_i とすると、次式のような不定性が残る。

$$\hat{W}_i = D_i P_i W_i \quad (1)$$

ここで、 $i = 1, 2, \dots, I$ は周波数ビンのインデックス、 W_i は分離行列、 $P_i \in \{0, 1\}^{N \times N}$ は分離行列 W_i の行ベクトル w_{in} の順番を入れ変えうるパーミュテーション行列 (置換行列)、 $D_i \in \mathbb{R}^{N \times N}$ は、 w_{in} のスケールを変化させる可能性のある対角行列を表す。

従って、FDICA で推定される分離信号

$$y_{ij} = \hat{W}_i x_{ij} \quad (2)$$

$$= [y_{ij1}, y_{ij2}, \dots, y_{ijn}, \dots, y_{ijN}]^T \in \mathbb{C}^N \quad (3)$$

は、推定音源の順番やスケールが周波数毎にばらばらになっている状態である。その概念図を Fig. 1 に示す。ここで、 y_{ij} は分離信号、 x_{ij} は観測信号、 $j = 1, 2, \dots, J$ は時間フレーム数を表す。FDICA

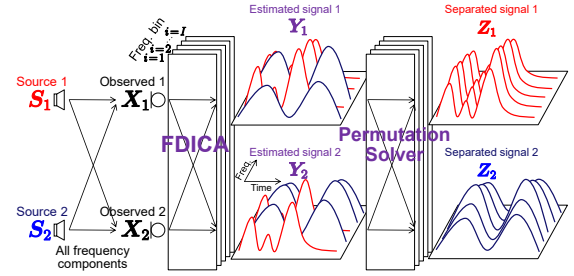


Fig. 1 Permutation problem in FDICA, where $N = M = 2$.

で得られる (パーミュテーション問題が生じている状態の) 推定信号 y_{ij} の n 番目のスペクトログラムを $Y_n \in \mathbb{C}^{I \times J}$ と定義する。このうち、 D_i によって生じるスケールの任意性は、プロジェクションバック法 [3] で解析的に復元可能である。一方で、 P_i によって生じる分離信号の順番の任意性 (パーミュテーション) を I 個の全周波数ビンに関して復元することは、組み合わせ爆発が生じるため容易ではない。このパーミュテーション問題を解決する処理は次式で表される。

$$z_{ij} = P_i^{-1} D_i^{-1} y_{ij} \quad (4)$$

従って、パーミュテーション問題の解決とは、全周波数ビンにわたって P_i^{-1} を求める問題として解釈できる。

3 提案手法

3.1 DNN の入出力

同一音源に属する成分の相関を強調するため、推定信号 (Y_1, Y_2) のパワースペクトラム ($|Y_1|^2, |Y_2|^2$) の比率に変換する正規化 [4] を施す。推定信号の正規化振幅スペクトログラム (\bar{Y}_1, \bar{Y}_2) から、時間フレーム j を中心とする局所時間振幅スペクトログラムを抽出し、一次元に整形 (ベクトル化) したものを DNN の入力とした。DNN による出力は次式で表される。

$$\hat{l}_j = \text{DNN}(d_j) \in [0, 1]^{2I} \quad (5)$$

ここで、 $\hat{l}_j = [\hat{l}_{11j}, \hat{l}_{21j}, \dots, \hat{l}_{1Ij}, \hat{l}_{2Ij}, \dots, \hat{l}_{2Ij}]^T$ は出力である予測ベクトルを表す。また DNN の出力には softmax 関数を用い、確率値とした。2 つのパーミュテーション問題が生じている入力信号

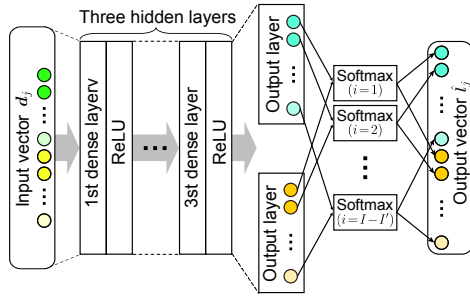


Fig. 2 DNN architecture.

($\tilde{Y}_{j1}, \tilde{Y}_{j2}$) の各周波数成分のそれぞれが「1 番目の音源の成分である確率 l_{i1} 」と「2 番目の音源の成分である確率 l_{i2} 」を入力ベクトルから予測したものと定義し、提案手法ではこの定義に基づいて正確な予測ができる DNN を学習する。

3.2 DNN の構造

Fig. 2 に提案深層パーミュテーション解決法で用いる DNN の構造を示す。DNN は入力層、隠れ層 3 層、及び出力層の計 5 層の全結合層からなる MLP となっている。

3.3 DNN 学習時の損失関数

損失関数には平均二乗誤差 (MSE) を用いる。具体的には、DNN の出力である予測ベクトルに従って、推定信号 (Y_1, Y_2) を並び替えを行い正解の分離信号の間で平均二乗誤差 (MSE) を使用する。

本文中に数式を挿入する場合は、次に示すように中央寄せすると良い。

$$a^2 + b^2 = c^2$$

複数行の数式を挿入する場合は、次のように等号記号で左右位置をそろえる。

$$\begin{aligned} a^2 + b^2 &= c^2 \\ d^2 &= e^2 + f^2 \\ A + B + C &= E + F + G \end{aligned}$$

横に長すぎる数式は改行を設けることで対処できる。但し、可能な限り左辺と右辺の左右位置は重ならないようにすることが望ましい。

$$\begin{aligned} A &= B + C + D + E + F + G + H + I + J + K \\ &\quad + L + M + N + O + P \end{aligned}$$

縦に幅を取る数式を挿入すると次のようになる。

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right)$$

なお、数式中に登場した変数等は全て登場直後に説明されなければならない。その一例を次式に示す。

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

ここで、 a は 2 乗項の係数、 b は 1 乗項の係数、 c は定数項をそれぞれ示す。本文全体で未定義の変数が残っていないか十分注意すること。

一度登場した数式を後の本文で引用する場合は、次式のように数式の右側に式番号を付与する。

$$A = \pi r^2 \quad (6)$$

式番号は括弧書きの数字とし、登場する順番に番号を付与する。式番号を用いて数式を引用する場合は「式 (6)」又は「(6) 式」と記述する。どちらを用いても良いが、表記は本文全体で統一されている必要がある。

文中 (インライン) に数式を挿入することもできる。例えば、 $f(x) = 1/2 \sin(\alpha + \beta)$ のように記述できる。ただし、縦に幅を取る数式はインラインで記述することは避けるべきである。また分数も、インラインで表記する場合は $V = 4/3\pi r^3$ のようにスラッシュ記号を用いて縦の幅を節約する工夫をする。

3.4 参考文献の引用の挿入方法

本文の適切な箇所で参考文献を引用する場合は、このように記述する [?]. この引用番号を表す記号は、予稿の末尾に記載されている参考文献リストと対応している [?]. 但し、登場する順に番号を付与する。なお、参考文献として不適切な文書を引用してはいけない。例えば、信憑性が疑われるインターネット上の記事 (ブログ、ウィキペディア、ニュース等) や学術的でない雑誌等は引用を避けるべきである。厳正な査読を経て掲載された原著論文及び国際会議論文が最も適切であるが、確立された技術を網羅した教科書等を引用する場合もある。その他、他者の学位論文 (修士論文、博士論文等) を引用することもあるが、これらの多くは信憑性が疑われるものも多々あるので十分注意する。

4 まとめ

以上の方法で作成すると、正しい体裁の技術文書を作成することができる。より読みやすく正しい技術文書を作成するためには、詳細を解説した参考書等を参考にすること。また、予稿完成後も複数人で念入りにチェックを繰り返し、誤字や脱字、不適切な記述、数式の誤り等が無いように十分注意すること。

5 注意事項

本テンプレートは元々 Word でのみしか配布されていなかった卒業研究発表会予稿のテンプレートを、北村大地が LaTeX による組版で可能な限り忠実に再現したファイルになります。マージンや余白などには細心の注意を払いましたが、ボールドフォントの違い等の細かい点は組版ソフトの違いが存在する以上生じてしまいます。あくまでも非公式のテンプレートとして自己責任で使用してください。

TeX のコンパイラは恐らく最もスタンダードである pdfLaTeX を使用することを想定しています。他のコンパイラでの動作確認はしておりません。ま

た、今後の対応の予定もありません。もう北村は疲れました。

参考文献

- [1] P. Comon, “Independent component analysis, a new concept?,” *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.
- [2] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain,” *Neurocomputing*, vol. 22, pp. 21–34, 1998.
- [3] K. Matsuoka and S. Nakashima, “Minimal distortion principle for blind source separation,” *Proc. ICA*, pp. 722–727, 2001.
- [4] H. Sawada, S. Araki, and S. Makino, “Measuring dependence of bin-wise separated signals for permutation alignment in frequency-domain BSS,” *Proc. ISCAS*, pp. 3247–3250, 2007.