

ユーザーからの補助情報を用いる 独立低ランク行列分析

大島風雅 (香川高専), 中野将生 (筑波大学), 北村大地 (香川高専)

1. 研究背景

• ブラインド音源分離: 混合系が不明な観測信号から混合前の音源を推定

$$x_{ij} = A_i s_{ij} \quad y_{ij} = W_i x_{ij}$$

混合行列 A_i 分離行列 W_i

$W_i = A_i^{-1}$ となるような分離行列 W_i を求める

i : 周波数インデクス j : 時間インデクス

• 非負値行列因子分解 (NMF): 低ランクモデリングによる行列分解

入力行列 $X_n(I \times J)$ 基底行列 $T_n(I \times K)$ アクティベーション行列 $V_n(K \times J)$

音源のスペクトルパターンとそのアクティベーションに分解

n : 音源インデクス

• 独立低ランク行列分析 (ILRMA): 周波数領域独立成分分析の音源モデルとして NMFを導入したブラインド音源分離

観測信号 \hat{x}_1, \hat{x}_2 → 周波数毎の分離行列 W_i → 分離信号 \hat{y}_1, \hat{y}_2 → 短時間フーリエ変換 → $|Y_1|, |Y_2|$ → NMFによる低ランクモデリング → $T_1 V_1, T_2 V_2$

各分離信号(音源)の時間周波数構造が低ランクな行列となるように周波数毎の分離フィルタを更新

ILRMAのコスト関数

$$\mathcal{J} = \sum_{i,j,n} \left[\frac{|w_{i,n}^H x_{ij}|^2}{\sum_k t_{ik,n} v_{kj,n}} + \log \sum_k t_{ik,n} v_{kj,n} \right] - 2J \sum_i \log |\det W_i|$$

$W_i = (w_{i,1}, \dots, w_{i,n}, \dots, w_{i,N})^H$ $t_{ik,n}: T_n$ の非負要素 $v_{kj,n}: V_n$ の非負要素

解決すべき問題

初期乱数によってブロックパーミュテーションが発生しILRMAの分離が失敗する

本研究の目的

ユーザーからのアノテーションを受けてブロックパーミュテーションを起こさないようにILRMAの最適化を誘導するアルゴリズムを改良, 実験的に調査, 考察する

2. 従来手法と提案手法

• インタラクティブ音源分離システム[中野ら, 2020]

ブロックパーミュテーション問題とは

- パーミュテーション不整合が**まとまった周波数**でおこること

ある音源の周波数帯域が別の音源に現れている

ユーザーが目視でブロックパーミュテーションを判別, ユーザーからのアノテーションをアルゴリズムに与える

アノテーション情報の活用法

- ブロックパーミュテーション発生周波数を直接指定
- 特定の音源のみが沈黙している時間を指定

• ユーザーからのアノテーション情報の活用法

- ブロックパーミュテーション発生周波数の指定

i_s, i_e : ブロックパーミュテーションが発生している周波数帯域 ($i_s < i_e$)

n_s : ブロックパーミュテーションが発生している音源

n_t : 移動先の音源インデクス

従来手法

$$w_{i_s, n_s}, \dots, w_{i_e, n_s} \Leftrightarrow w_{i_s, n_t}, \dots, w_{i_e, n_t}$$

$$t_{i_s k, n_s}, \dots, t_{i_e k, n_s} \Leftrightarrow t_{i_s k, n_t}, \dots, t_{i_e k, n_t} \quad \forall k$$

提案手法 **追加**

$$v_{kj, n} \leftarrow \rho \quad \forall k, j, n$$

$\rho: (0, 1)$ の一様乱数

ねらい

アクティベーション行列をリセットすることで局所解から抜け出すことを期待

2. 従来手法と提案手法(続き)

- 特定の音源 n_t のみが沈黙している時間の指定

従来手法

$$v_{kj_s, n_t}, \dots, v_{kj_e, n_t} \leftarrow \epsilon v_{kj_s, n_t}, \dots, \epsilon v_{kj_e, n_t} \quad \forall k$$

提案手法(a) **変更**

$$w_{i, n} \leftarrow \rho \quad \forall i, n$$

提案手法(b) **追加** ϵ : 正の微小値 (10^{-15})

$$v_{k1, n_t}, \dots, v_{k(j_s-1), n_t} \leftarrow \alpha, \alpha, \dots, \alpha \quad \forall k$$

$$v_{k(j_s+1), n_t}, \dots, v_{kJ, n_t} \leftarrow \alpha, \alpha, \dots, \alpha \quad \forall k$$

$$v_{kj, n} \leftarrow \alpha \quad \forall k, j, n \neq n_t$$

$\alpha: [1.0 \times 10^5, 1.1 \times 10^5]$ の乱数

ねらい

周波数のときと同様に, パラメタをリセットすることで局所解から抜け出せることを期待

- なぜ特定の音源が沈黙している時間を指定するか

反復射影法[Ono, 2011]により分離行列の更新式は次式で表される

$$U_{i,n} = \frac{1}{J} \sum_j \frac{1}{\sum_k t_{ik,n} v_{kj,n}} x_{ij} x_{ij}^H$$

$$w_{i,n} \leftarrow (W_i U_{i,n})^{-1} e_n$$

下線部が0に近づく

$$w_{i,n} \leftarrow w_{i,n} (w_{i,n}^H U_{i,n} w_{i,n})^{-\frac{1}{2}}$$

特定の音源だけが沈黙している時間を指定することで, それ以外の音源間の周波数毎の空間相関 $U_{i,n}$ がより正確に求まる

提案手法(a), (b) 共通

提案手法(a)

提案手法(b)

3. 比較実験

• 実験条件

実験信号 (SiSEC2011のUNDタスク)

混合信号	音源
no. 1	dev1_female3_synthconv_130ms_5cm_sim_1
	dev1_female3_synthconv_130ms_5cm_sim_2
no. 2	dev1_male3_synthconv_130ms_5cm_sim_1
	dev1_male3_synthconv_130ms_5cm_sim_2
no. 3	dev1_male3_synthconv_130ms_5cm_sim_1
	dev1_female3_synthconv_130ms_5cm_sim_2

実験パラメタ

パラメタ	値
STFT長	128ms
STFT窓関数	ハミング窓
NMF基底数	3
ILRMA反復回数	160

80回ILRMAの反復を行った後, アノテーションを与えた場合と与えなかった場合のSDR (source-to-distortion ratio) を比較

• 実験結果

- 周波数アノテーションを与えた場合 (a) no.1, (b) no.2, (c) no.3
- 時間アノテーションを与えた場合 (a) no.1, (b) no.2, (c) no.3

Legend: Simple ILRMA (dashed), ILRMA w/ conventional freq. annotation (blue), ILRMA w/ proposed freq. annotation (red), ILRMA w/ conventional time annotation (green), ILRMA w/ proposed time annotation (a) (red), ILRMA w/ proposed time annotation (b) (green).

全ての信号において提案手法は従来手法を上回るか, 同等の精度を示している
提案手法は従来手法と比較して**安定かつ高精度**

原稿の正誤表

p.2 2.3.1節 l.1	誤: $i = e$ 正: $i = i_e$
表記揺れの修正	NMF基底数 誤: L 正: K
	NMF基底インデクス 誤: l 正: k