

Bayesian NMF2D による音楽音響信号解析

福田 鉄也[†] 久保田 光一[†]

[†] 中央大学大学院 理工学研究科 情報工学専攻

1 はじめに

近年、情報技術の発展に伴って大量の音楽データに対する楽曲の内容に基づいた検索技術などへの需要が高まっている。このような技術の基盤として近年注目されているものの1つに、非負値行列分解 (Nonnegative Matrix Factorization; NMF) と呼ばれる手法に基づいて、音楽音響信号を構成している特徴的なパターンを抽出する技術がある。

一方で NMF はその柔軟性から微細なノイズを1つの音響パターンとして捉えるなど、しばしば人間にとっては解釈しづらい結果を抽出することがあった。そこで本稿では NMF の拡張の1つであり、Nonnegative Matrix Factor 2D Deconvolution (NMF2D) と呼ばれる音楽音響信号をいくつかのパーツの対数周波数・時間平面上での2次元畳み込みで表現したモデルを元にしたベイズモデルの構築法とその推論法について述べる。また、実際の音楽音響信号に対して適用した場合の実験結果について評価する。

2 NMF2D

NMF [3, 4] は、以下のように任意の非負値行列 $Y \in \mathbb{R}^{\geq 0, M \times N}$ を2つの非負値行列 $W \in \mathbb{R}^{\geq 0, M \times K}$ および $H \in \mathbb{R}^{\geq 0, K \times N}$ の積で近似する手法である。

$$Y \approx WH \quad (1)$$

音楽音響信号に対して NMF を適用する場合、観測された周波数ビン数 M 、フレーム数 N の振幅スペクトログラムやウェーブレットスペクトログラムを $M \times N$ の非負値行列 Y とみなし、その周波数ビン m 、フレーム n における成分 $Y_{m,n}$ が K 個の要素の結合で表現されると仮定するものが多い [7]。

また、NMF を拡張し、観測スペクトログラムを以下のように時間に依存するパラメータ τ を持つ W^τ と、音程に依存するパラメータ ϕ を持つ H^ϕ の2次元畳み込みで表現したものを NMF2D [5, 6] という。

$$Y \approx X = \sum_{\tau=0}^{T-1} \sum_{\phi=0}^{F-1} \overset{\downarrow \phi}{W}^\tau H^\phi \quad (2)$$

ここで、 $\overset{\downarrow n}{A}$ は行列 A の各行を下に n シフトする操作、

Bayesian nonnegative matrix factor 2D deconvolution for music signal analysis

Tetsuya FUKUDA[†], Koichi KUBOTA[†]

[†]Information and System Engineering Course, Graduate School of Science and Engineering, CHUO University

$\overset{\rightarrow n}{A}$ は行列 A の各列を右に n シフトする操作である。例えば

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \quad (3)$$

のとき、

$$\overset{\downarrow 1}{A} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \overset{\rightarrow 2}{A} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 4 \\ 0 & 0 & 7 \end{pmatrix} \quad (4)$$

のようになる。

NMF2D は $F = 1$ としたときに Nonnegative Matrix Factor Deconvolution (NMFD) [8] と、 $F = 1$ かつ $T = 1$ としたときに NMF と等価なモデルとなっている。一方で NMF2D においてはそれぞれのスペクトルパーツの時間、周波数軸上でのシフト不変性が仮定されているため、周波数軸として調波音の周波数の相対位置が基本周波数に依存しない対数周波数が用いられることが多い。

3 Bayesian NMF2D

NMF や NMF2D は一般的に、観測行列と再構成行列の間に何らかの「乖離度」を設計し、これを最小化するような最適化問題として定式化される。この乖離度の設計に良く利用される規準として二乗誤差、一般化 Kullback-Leibler ダイバージェンス [4]、板倉斎藤擬距離 [2] などが挙げられる。

また、一般化 Kullback-Leibler ダイバージェンス規準の NMF は、観測行列の各行列要素 $Y_{m,n}$ が再構成行列の各行列要素をパラメータに持つ Poisson 分布に従って独立に生成されたと仮定したときの最尤推定問題と等価であることが知られている [1]。

ここで、NMF2D における再構成行列 X の各行列要素 $X_{m,n}$ は以下のように書き下すことができる。

$$X_{m,n} = \sum_{k=1}^K \sum_{\tau=0}^{T-1} \sum_{\phi=0}^{F-1} W_{m-\phi,k}^\tau H_{k,n-\tau}^\phi \quad (5)$$

したがって一般化 Kullback-Leibler ダイバージェンス規準の NMF2D について以下のようなモデルを考えることができる。

$$Y_{m,n} \sim \text{Poisson} \left(\sum_k \sum_\tau \sum_\phi W_{m-\phi,k}^\tau H_{k,n-\tau}^\phi \right) \quad (6)$$

推論の便宜上、 $W_{m,k}^\tau$ と $H_{k,n}^\phi$ の事前分布については Poisson 分布の共役事前分布であるガンマ分布を設定

する.

$$W_{m,k}^\tau \sim \text{Gamma}(a_W, b_W) \quad (7)$$

$$H_{k,n}^\phi \sim \text{Gamma}(a_H, b_H) \quad (8)$$

ここで, a_W, b_W および a_H, b_H は事前分布を制御するハイパーパラメータである.

4 変分ベイズ法

いま観測データ Y が与えられたとき, 確率変数の事後分布 $p(W, H|Y)$ を求めたいが, 真の事後分布を解析的に計算することは容易ではない. そこで, 以下のような分解が可能な関数形を持つ分布 $q(W, H)$ を考え, $p(W, H|Y)$ との間の Kullback-Leibler ダイバージェンスが最小になるように反復更新することで真の事後分布を近似的に計算する.

$$q(W, H) = \prod_{m,k,\tau} q(W_{m,k}^\tau) \prod_{k,n,\phi} q(H_{k,n}^\phi) \quad (9)$$

これは対数周辺尤度 $\log p(Y)$ の変分下界 \mathcal{L} を最大化することと等価である. ここで \mathcal{L} も解析的に求めることはできないが, 対数関数が凹関数であることから Jensen の不等式を用いて \mathcal{L} の更なる下界を設計し, これについての最大化を考えることができる. スペースの都合上詳細は省略するが, Bayesian NMF [1] とよく似た閉形式での更新則が導出できる. また, 行列計算の形での効率的な実装も可能である.

5 実験

提案法の動作を確認するための実験を行った. 観測スペクトログラムは Gabor ウェーブレット変換 (サンプリング周波数 16kHz, 時間分解能 10ms, 周波数分解能 25cent, 最低周波数 55Hz) によって計算したものをを用いた. またハイパーパラメータはそれぞれ $a_W = b_W = 1$, $a_H = b_H = 1$ に設定した.

$K = 4, F = 48, T = 20$ としてトランペットとピアノの混合音に対して適用した結果を図 1 に示す. 2つのコンポーネントがそれぞれピアノとトランペットのスペクトルに対応した. また, 4個のコンポーネントのうちこの2つを除いて $W_{m,k}^\tau, H_{k,n}^\phi$ とともにほぼ消滅した. このように小さな値をもつ成分がより消滅しやすくなる効果はベイズ推定に基づくものであり, パラメータを点推定する最尤推定や MAP 推定では得ることができない. また, a_W, a_H が小さいほどこの効果は大きくなる.

6 おわりに

本稿では NMF2D と呼ばれるモデルのベイズ拡張について述べた. Bayesian NMF2D では一部のコンポーネントがそれぞれの楽器音に対応し, 観測スペクトログラムの説明に不要なコンポーネントは自動的に消滅する傾向があった.

F や T などのパラメータの取り扱いや, 一般化 Kullback-Leibler ダイバージェンス規準よりも性能が

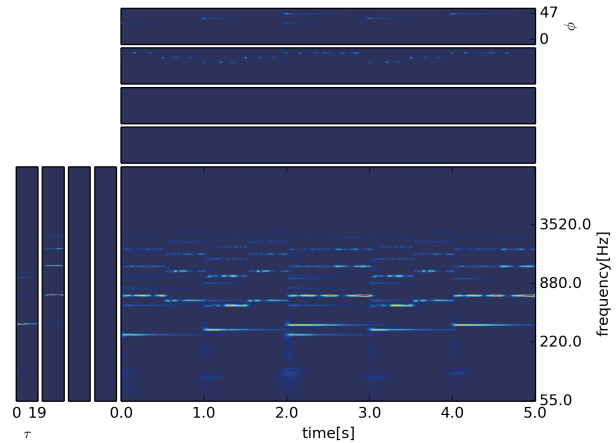


図 1 トランペットとピアノの混合音に対して適用した場合の観測スペクトログラム Y (右下) と推定された W^τ (左), H^ϕ (上).

良いとされる二乗誤差規準の NMF2D に対するベイズモデルの構築と効率的な推論法の導出は今後の課題である.

参考文献

- [1] A. T. Cemgil, "Bayesian Inference in Non-negative Matrix Factorisation Models," *Computational Intelligence and Neuroscience*, 2009.
- [2] C. Févotte, N. Bertin and J. -L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural Computation*, 21(3), pp.793–830, 2009.
- [3] D. D. Lee and H. S. Seung, "Learning the parts of objects with nonnegative matrix factorization," *Nature*, 401, pp.788–791, 1999.
- [4] D. D. Lee and H. S. Seung, "Algorithms for nonnegative matrix factorization," in *Adv. NIPS*, pp.556–562, 2000.
- [5] M. Mørup and M. N. Schmidt, "Sparse non-negative matrix factor 2-D deconvolution," Technical Report, Technical University of Denmark, 2006.
- [6] M. N. Schmidt and M. Mørup, "Nonnegative matrix factor 2-D deconvolution for blind single channel source separation," in *Proc. ICA2006*, pp.700–707, 2006.
- [7] P. Smaragdis and J. C. Brown, "Non-Negative Matrix Factorization for Polyphonic Music Transcription," in *Proc. WASPAA*, pp.177–180, 2003.
- [8] P. Smaragdis, "Non-negative Matrix Factor Deconvolution; Extracation of Multiple Sound Sources from Monophonic Inputs," in *Proc. ICA2004*, pp.494–499, 2004.

付録 A 確率分布の表記

$$\text{Poisson}(x|\lambda) = \frac{\lambda^x}{\Gamma(x+1)} e^{-\lambda}$$

$$\text{Gamma}(x|a, b) = \frac{1}{\Gamma(a)} b^a x^{a-1} e^{-bx}$$