

令和 6 年 6 月 16 日現在

機関番号：32682

研究種目：若手研究

研究期間：2021～2023

課題番号：21K17769

研究課題名（和文）ニューラルネットワークを用いた統計モデルの学習と音響信号処理への応用

研究課題名（英文）Learning statistical models using neural networks and application to acoustic signal processing

研究代表者

丹治 寛樹 (Tanji, Hiroki)

明治大学・研究・知財戦略機構（生田）・研究推進員（ポスト・ドクター）

研究者番号：40896255

交付決定額（研究期間全体）：（直接経費） 2,900,000円

研究成果の概要（和文）：本研究では、信号分離および雑音除去のための統計モデルを学習する手法の確立を目標とする。そのために、複素スペクトログラムに対する確率密度関数を自動的に学習できるか、という観点に基づき、ニューラルネットワークを用いた確率密度関数および最適化アルゴリズムの構築とその学習方法を提案した。特に非負値行列因子分解およびその拡張における統計モデルに着目し、信号分離および雑音除去のタスクにおけるニューラルネットワークの学習手法を提案した。

研究成果の学術的意義や社会的意義

統計モデルに基づくNMFの乖離度は一般に信号分離や雑音除去といったタスクにおける性能評価指標を考慮していないため、これらのタスクにおいて常に高い性能を発揮するとは限らなかった。しかも、従来の乖離度の調整可能なパラメータはせいぜい1つしか無く、モデルの表現力は限られていた。一方では、提案法ではニューラルネットワーク（NN）を用いて乖離度を構成することでモデルの表現力を向上させ、評価指標を考慮してNNを学習することに成功した。また、シミュレーションにより、信号分離や雑音除去においてNMFの乖離度が与える性能への影響を明らかにした。これはNMFの乖離度に関する研究を総括する意味で有意義である。

研究成果の概要（英文）：The goal of this study is to establish a method for learning statistical models for signal separation and noise reduction. For this purpose, we proposed a neural network-based probability density function and optimization algorithm and its learning method, based on the viewpoint of whether the probability density function for complex spectrograms can be automatically learned. In particular, we focused on statistical models for non-negative matrix factorization and its extensions and proposed neural network learning methods for signal separation and noise reduction tasks.

研究分野：知覚情報処理関連

キーワード：非負値行列因子分解 音源分離 雑音除去 深層学習 深層展開 統計モデル

様式 C-19、F-19-1 (共通)

1. 研究開始当初の背景

非負値行列因子分解 (Nonnegative Matrix Factorization; NMF) は、全ての要素が非負の観測行列を観測行列よりも低ランクな 2 つの非負行列の積によって近似する枠組みである。例えば、音響信号処理においては、パワースペクトログラムが観測行列としてよく用いられる。図 1 に NMF によるパワースペクトログラムの分解例を示す。パワースペクトログラムは、横軸を時間、縦軸を周波数とし、各要素の値は、ある時間におけるある周波数成分

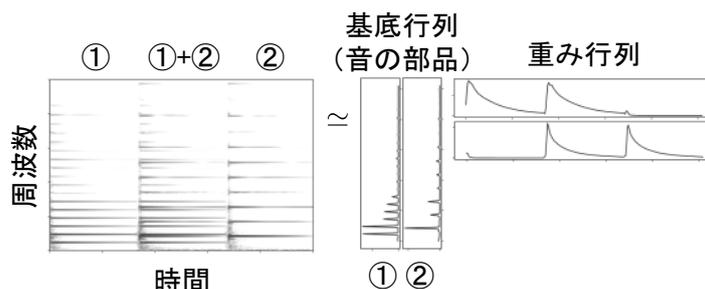


図1：NMFによるパワースペクトログラムの分解例

の信号の強度を意味する。NMF をパワースペクトログラムに適用することにより、パワースペクトログラムに頻繁に表れるスペクトルの集合、すなわち、「音の部品」から構成される行列 (**基底行列**) と、各部品の強度の時間変化を表す行列 (**重み行列**) に分解できる。したがって、NMF により、パワースペクトログラムは、「音の部品」の足し合わせによって近似される。このような音響信号における NMF の作用は、混ざり合った音を分離する「音源分離」や、雑音に埋もれた音を復元する「雑音除去」のタスクにおいて、観測信号または音源信号の近似を得るために非常に有用である。したがって、音響信号処理において NMF は基礎的な技術の一つとしてよく知られている。

NMF の分解表現は、基底行列と重み行列の積によって近似された行列と観測行列の乖離度を最小にすることで得られる。乖離度は近似の「良さ」を評価する指標であるため、最小化する乖離度によって得られる分解表現が変わる。そこで、本研究では、**NMF の乖離度に着目**する。NMF の研究の黎明期では、観測された非負行列の各要素は指数分布族を用いてモデリングされ、指数分布族に基づく乖離度として Bregman divergence がよく用いられてきた。音響信号処理において、NMF はパワースペクトログラムを分解する枠組みである一方で、観測行列であるスペクトログラムは、パワーに加えて位相の情報を持った複素数で与えられることに注意する必要がある。そこで、**複素数のスペクトログラムを分解する枠組みとして NMF を定式化**できないだろうか、という観点で音響信号処理の分野で NMF の乖離度に関する研究が発展した。

Fevotte らは、複素正規分布を用いた統計モデルに基づき、NMF の乖離度を導出した。この統計モデルでは、スペクトログラムの各要素が 1 変量の複素正規分布に従うと仮定する。この仮定によると、**複素正規分布の分散をパワースペクトログラムの推定値とみなす**ことができる。しかし、スペクトログラムの従う分布として必ずしも**複素正規分布にこだわる必要はない**。そこで、信号分離や雑音除去などにおける性能を改善するために、これまでに Fevotte らの統計モデルに替わるモデルや複数のモデルを含む一般化が検討されてきた。これらの従来研究では、まず、ある特定の複素分布に着目し、その**確率密度関数から NMF の乖離度を導出**する。この乖離度は、しばしば調整可能なパラメータを持つ。次に、この乖離度を最小化するための反復更新アルゴリズムを乖離度の勾配に基づいて導出する。

アルゴリズムの性能は、NMF を信号分離や雑音除去に適用することで評価される。しかし、**従来研究のどの乖離度も性能評価に用いられる指標を考慮していない**ため、常に高い性能を発揮するとは限らない。また、評価に用いるタスクや音源信号など、**乖離度やその調整可能なパラメータ以外の要因に実験結果が左右**される。これらのことから、NMF の乖離度が性能に与える影響や、性能を向上させるためのパラメータの調整方法は明らかになっていない。

2. 研究の目的

本研究では、「**NMF の統計モデルにおいて、複素スペクトログラムおよびパワースペクトログラムに対する確率密度関数を自動的に学習できるか**」という問いを提起する。この問いに答えるために、本研究ではニューラルネットワークが任意の関数を表現できることに着目し、確率密度関数または統計モデルとして解釈可能なニューラルネットワークの構築と学習方法の確率を目的とする。本研究では、NMF に加えて、音源分離のための NMF の拡張モデルも研究対象とする。本研究では、以下の 3 点を明らかにする。

- (1) ニューラルネットワークにより生成した確率密度関数に基づく NMF の乖離度、および、基底行列と重み行列の推定アルゴリズム
- (2) NMF を用いた信号分離および雑音除去におけるニューラルネットワークの学習方法
- (3) NMF の多チャンネル拡張を用いた信号分離におけるニューラルネットワークの学習方法

3. 研究の方法

本研究において **NMF の乖離度はニューラルネットワークにより構築**される。したがって、この乖離度を基底および重みについて最小化するための **アルゴリズムも同じニューラルネット**

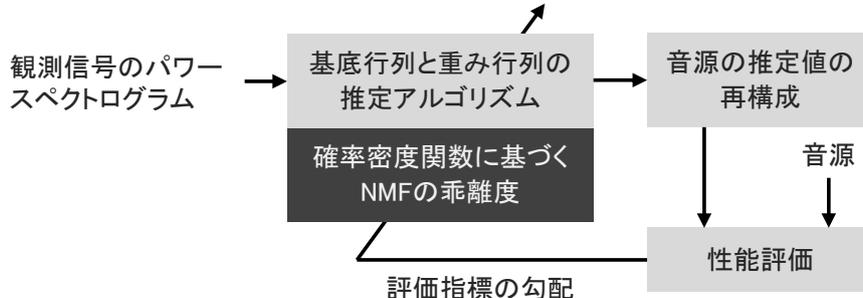


図2：提案法の模式図

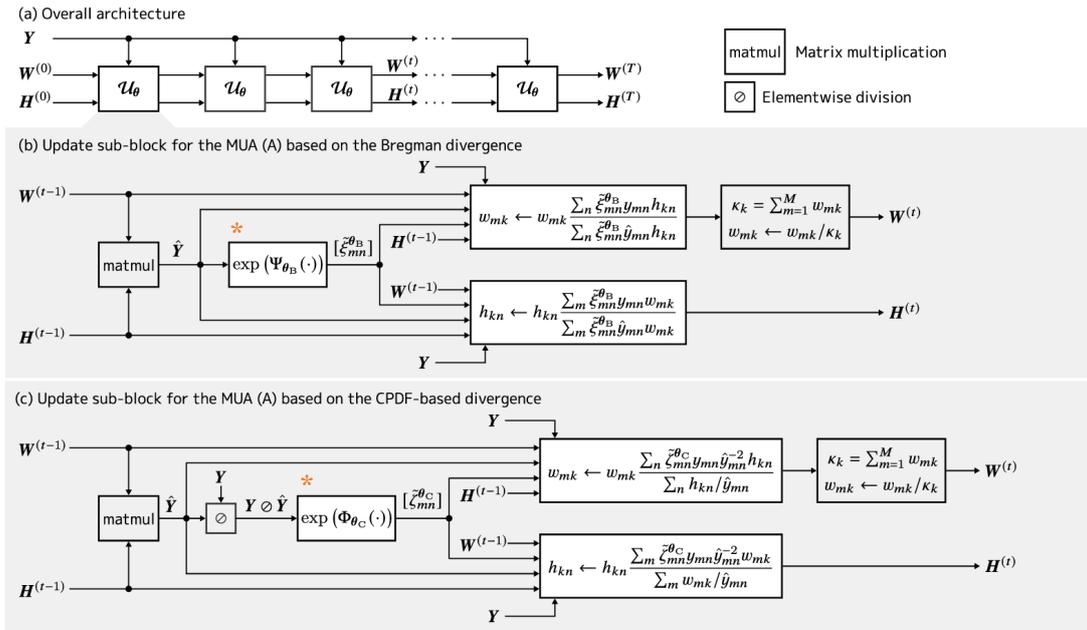


図3：提案したネットワーク構造

ワークを用いて表現できる。本研究では、複素確率密度関数に基づく乖離度 (CPDF-based divergence) および Bregman divergence に着目し、それぞれの乖離度に基づく評価関数に対して、評価関数を発見的に最小化するアルゴリズムを導出する。さらに、導出したアルゴリズムの乖離度由来する項をニューラルネットワークの出力を用いて置き換えることで、最小化アルゴリズムおよび乖離度の両方として解釈可能なニューラルネットワークを構築する。

図2に提案法の模式図を示す。提案するニューラルネットワークは、従来のNMFの最適化アルゴリズムと同様に観測信号のパワースペクトログラムが入力され、基底および重みの推定値が出力される。さらに、雑音除去や信号分離といったタスクに応じて音源の推定値を再構成する。雑音除去や信号分離における性能を向上させる乖離度を見つけるために、音源の真値と推定値を用いてそれぞれのタスクにおける評価指標を算出し、評価指標の勾配に基づいてニューラルネットワークを学習する。

4. 研究成果

図3に提案したネットワーク構造 (deep multiplicative update algorithm; DeMUA) を示す。図2(a)において、DeMUAは、観測されたスペクトログラムと基底および重みの初期値を受け取り、パラメータの更新サブブロックを経由して基底および重みの推定値を出力する。図2(b)(c)はそれぞれ Bregman divergence および CPDF-based divergence の更新サブブロックである。図中に*で示したブロックは NMF の乖離度由来する関数であり、学習パラメータを持つニューラルネットワークで構築する。図2(b)(c)は、直接的には NMF のパラメータの更新則を表現しているが、*で示したブロックの関数についての微分方程式を解くことで NMF の乖離度を取り出すことができる。

本研究では、図2に示した DeMUA をスパース雑音除去および教師あり信号分離に適用し、シミュレーションにより評価した。

スパース雑音除去のシミュレーションにはエレクトリックギターにより演奏された12個の音源信号を用い、そのうちの6個 (SID1-1) を学習用データ、残りの6個 (SID1-2) をテスト用データとした。時間周波数領域でスパースなノイズを人工的に合成し、音源信号に重畳することで観測信号とした。シミュレーションでは、ニューラルネットワークを用いて構成した乖離度に

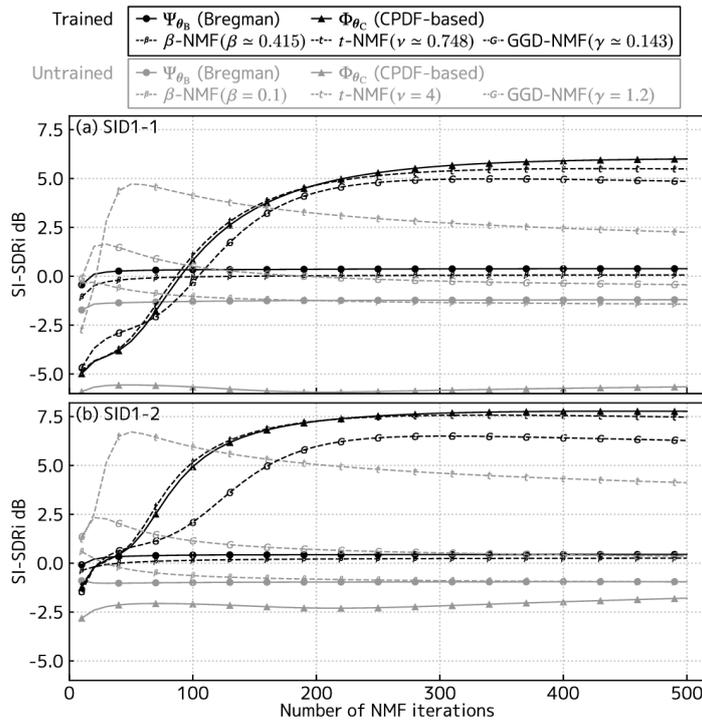


図4：スパース雑音除去における提案法の評価結果

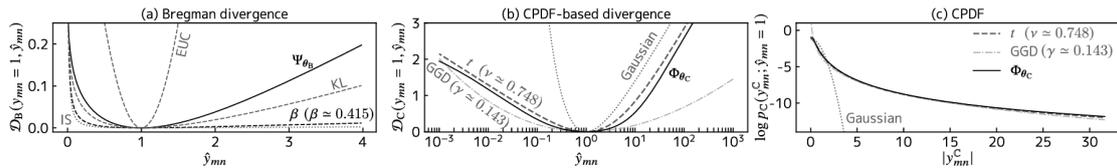


図5：提案法により得られた乖離度(a)(b)および確率密度関数(c)

加えて、従来のハイパーパラメータを持つ確率密度関数に基づく NMF (t-NMF, GGD-NMF) も提案法と同じ条件で学習した. 雑音除去の性能は、音源信号との誤差のパワーを評価する scale-invariant signal-to-distortion ratio (SI-SDR) の改善量 SI-SDR_i を用いて評価した.

図 4 に NMF のアルゴリズムによる SI-SDR_i の推移を示す. 図 4 より、全てのネットワークが未学習時と比較して SI-SDR を向上させていることがわかる. 特に CPDF-based divergence に基づく NMF の SI-SDR_i が最も高い. 図 4 では、未学習の t-NMF および GGD-NMF の SI-SDR は約 50 回の更新後に劣化するが、提案法によりハイパーパラメータを学習することでこの性能の劣化を防げることがわかる.

図 5 に DeMUA を学習することで得られた乖離度および確率密度関数を示す. 図 5(b) に示した乖離度は観測データよりも推定値が小さな値となることに寛容であり、これにより推定値が突発的なノイズに影響されづらくなっていると考えられる. また、ニューラルネットワークにより構築した CPDF-based divergence と t-NMF および GGD-NMF の乖離度は似通っており、図 4 でこれらの NMF の性能が比較的高いことを支持している.

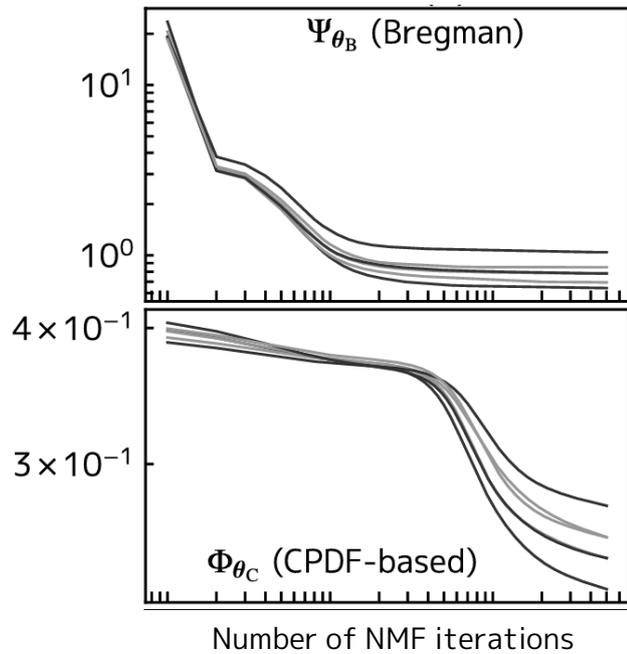


図6：NMFの評価関数の推移

図 6 に学習後のニューラルネットワークから算出した評価関数の推移を示す。DeMUA で用いている最適化アルゴリズムでは評価関数の単調減少は保証されていないが、図 6 では、全ての例において評価関数が単調減少していることがわかる。

スパース雑音除去に加え、教師あり信号分離に対しても提案法を適用し、提案法により乖離度を学習することで複数のデータセットに汎化性能をもつ乖離度を学習できることを示した。

NMF の多チャンネル拡張を用いた信号分離アルゴリズムの統計モデルを学習することを目的とし、上記の知見をもとに、信号分離のアルゴリズムにニューラルネットワークを組み込むことを試みた。この研究では、信号分離のアルゴリズムに現れる確率変数の期待値の項が音源信号の真値を含むように意図的に統計モデルを構築し、確率変数の期待値の項を雑音除去のための汎用的なニューラルネットワークで置換できることを突き止めた。当初予定していたエフォートを確保できず、効果的なニューラルネットワークの学習方法を発見するに至らなかった。しかし、音源信号の真値を用いたシミュレーションでは、3つの音源信号を2つのマイクロホンで録音し分離するという難易度の高いタスクにおいて、新たに導出したアルゴリズムが従来の NMF の拡張アルゴリズムよりも高い分離性能を持つことがわかった。この知見は、今後のニューラルネットワークを用いた NMF の多チャンネル拡張の研究において、分離性能を向上させるための足がかりになると考える。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Hiroki Tanji, Takahiro Murakami	4. 巻 E106-A
2. 論文標題 Deep Multiplicative Update Algorithm for Nonnegative Matrix Factorization and Its Application to Audio Signals	5. 発行年 2023年
3. 雑誌名 IEICE Trans. Fundamentals	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計6件（うち招待講演 0件/うち国際学会 2件）

1. 発表者名 Masaki Ikeuchi, Hiroki Tanji, and Takahiro Murakami
2. 発表標題 Improvement of the direction-of-arrival estimation method using a single channel microphone by correcting a spectral slope of speech
3. 学会等名 Proc. 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) (国際学会)
4. 発表年 2022年

1. 発表者名 Hiroki Tanji, Takahiro Murakami
2. 発表標題 Learning the statistical model of the NMF using the deep multiplicative update algorithm with applications
3. 学会等名 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2021 (APSIPA ASC 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 池内 証, 小野寺章吾, 丹治寛樹, 村上隆啓
2. 発表標題 最尤推定に基づく1ch 到来方向推定法の音声信号への拡張
3. 学会等名 電気学会システム研究会
4. 発表年 2021年

1. 発表者名 小野寺章吾, 池内柁, 丹治寛樹, 村上隆啓
2. 発表標題 到来方向に誤差を含む参照信号を利用したマイクロホンアレーの校正
3. 学会等名 電気学会システム研究会
4. 発表年 2021年

1. 発表者名 八木澤太貴, 丹治寛樹, 村上隆啓
2. 発表標題 Phase Vocoder を用いた話速変換におけるTransient Smearingの軽減
3. 学会等名 第38回信号処理シンポジウム
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関