4V-03

von Mises - Bernoulli DNN を用いた音源定位の検討

正木俊伍 [1], 小島諒介 [2], 杉山治 [3], 中臺一博 [1,4], 糸山克寿 [1], 西田健次 [1]
1 東京工業大学工学院システム制御系 2 京都大学医学研究科人間健康科学系
3 京都大学先制医療・生活習慣病研究センター 4 ホンダ・リサーチ・インスティチュート・ジャパン

1 はじめに

音源定位はマイクロホンアレイを用いた音響信号処理における重要なタスクの一つであり，マイクロホン間の位相差や強度差に基づいて音源の方向や位置を推定するものである．音源定位には multiple signal classification (MUSIC) 法 [1] を用いることが一般的であるが，計算コストの大きさなどの課題があるため，代替アプローチやこれを拡張した手法 [2] が研究されているが，その性能は伝達関数に依存してしまう．この伝達関数を深層学習で学習できれば音源定位の性能向上が期待できる．深層学習は音声認識 [3] を始め，ニューラルネットワークを用いて，高精度な識別学習，回帰学習ができる枠組みとして近年広く用いられている．音声認識では，振幅スペクトルなどの実数値をニューラルネットワークへ入力することが一般的であるため，あまり問題とならない．しかし音源定位では，2 次元の量である複素スペクトル（複素数）や周期性を持つ量である位相が定位のキューとなるため，これらをニューラルネットワークの入力として用いることが適当であるが，これらをそのまま入力してもネットワークが適切に学習されない．関連研究でも，振幅スペクトルを入力する手法 [4] ，複素スペクトルを前処理により実数に変換して入力する手法 [5] などが提案されているが，位相を直接ネットワークに入力する手法はほとんど存在しない．

本稿では，位相を直接入力することのできるニューラルネットワーク von Mises-Bernoulli deep neural network (vM-B DNN) を用いた音源定位手法を提案する．vM-B DNN は，周期的な量に対してよく用いられる von Mises 分布を用いて位相を直接入力することができるように拡張したニューラルネットワークである．シミュレーション実験により，vM-B DNN が位相を入力とした音源定位に利用可能であることを示す．

2 von Mises - Bernoulli DNN (vM-B DNN)

sigmoid 関数を用いたフィードフォワードニューラルネットワークの各層の計算は RBM(restricted Boltzmann machine) の隠れ層に対する事後分布の計算と一致することから，ニューラルネットワークの隠れ層の計算は RBM の隠れ層の点推定とみなすことができる．vM-B DNN は，vM-B RBM を考え，その点推定を基にした手法である．

2.1 von Mises - Bernoulli RBM (vM-B RBM)

RBM とは, 可視層と隠れ層間のノードの繋がり方が制限されている確率モデルである．入力および出力のノードの状態をそれぞれ $\boldsymbol{v} \in \{0,1\}^I$,$\boldsymbol{h} \in \{0,1\}^J$ としたとき, RBM の確率モデル $P(\boldsymbol{v},\boldsymbol{h})$ は (1) 式のように定義される．

$$P(\boldsymbol{v},\boldsymbol{h}) = -\frac{E(\boldsymbol{v},\boldsymbol{h})}{Z} \tag{1}$$

$$E(\boldsymbol{v},\boldsymbol{h}) = -\boldsymbol{a}^T\boldsymbol{v} - \boldsymbol{b}^T\boldsymbol{h} - \boldsymbol{v}^T\boldsymbol{W}\boldsymbol{h} \tag{2}$$

ここで $\boldsymbol{a},\boldsymbol{b},\boldsymbol{W}$ はパラメータ, Z は正規化定数である．このとき, 入力と出力の条件付確率 $P(v_i|\boldsymbol{h})$,$P(h_j|\boldsymbol{v})$ はともに Bernoulli 分布に従う．そのため, 上式によって定義された RBM は Bernoulli - Bernoulli RBM (B-B RBM) と呼ばれる．

一方, 今回提案するモデルである vM-B RBM では, 入力が von Mises 分布, 出力が Bernoulli 分布に従う信号であると仮定した際の RBM であり, 指数分布に関する RBM [6] の特殊な場合である．von Mises 分布 $vM(\cdot)$ [7] は, 確率変数 $\theta \in [0, 2\pi)$ を用いて (3) 式で表される分布である．

$$vM(\theta;\mu,\beta) = \frac{\exp(\beta\cos(\theta-\mu))}{I_0(\beta)} \tag{3}$$

ここで μ は平均方向, β は集中度を表すパラメータ, $I_0(\cdot)$ は第一種変形ベッセル関数である．

vM-B RBM は, $P(v_i|\boldsymbol{h})$ が von Mises 分布, $P(h_j|\boldsymbol{v})$ が Bernoulli 分布に従うように, $E(\boldsymbol{v},\boldsymbol{h})$ を (4) 式で定義する．

$$\begin{aligned} E(\boldsymbol{v},\boldsymbol{h}) = &-\boldsymbol{a}^T\cos(\boldsymbol{v}) - \boldsymbol{b}^T\sin(\boldsymbol{v}) \\ &- \boldsymbol{c}^T\boldsymbol{h} - (\cos(\boldsymbol{v})^T\boldsymbol{W} + \sin(\boldsymbol{v})^T\boldsymbol{Q})\boldsymbol{h} \end{aligned} \tag{4}$$

ここで $\boldsymbol{v} \in [0,2\pi)^I, \boldsymbol{h} \in \{0,1\}^J$ であり, $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{W},\boldsymbol{Q}$ はパラメータである．

2.2 von Mises - Bernoulli DNN の構築

DNN の各層が RBM の隠れ層の点推定とみなせることから，同様に，vM-B RBM の点推定を考えることで，vM-B DNN の計算を定義することができる．vM-B DNN の入力層の出力 $P(h_j = 1|\boldsymbol{v})$ を以下のように定義する．

$$P(h_j = 1|\boldsymbol{v}) = \frac{1}{1+\exp(-\hat{c_j})} \tag{5}$$

$$\hat{c_j} = c_j + \sum_i \cos(v_i)W_{ij} + \sin(v_i)Q_{ij} \tag{6}$$

ここで (5) 式は sigmoid 関数であり, (6) 式の $\hat{c_j}$ が B-B RBM と異なるだけである．ゆえに入力層は, sigmoid 関数を活性化関数とするニューラルネットワークと比べて $\hat{c_j}$ を変更するだけで実装できる．また，vM-B DNN では 2 層目以降は通常のニューラルネットワークと同様に構築する．

Examination of sound source localization with von Mises - Bernoulli DNN

Shungo Masaki[1], Ryosuke Kojima[2], Osamu Sugiyama[3], Kazuhiro Nakadai[1,4], Katsutoshi Itoyama[1], Kenji Nishida[1]
1 Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology
2 Graduate School of Medicine and Faculty of Medicine Kyoto University
3 Preemptive Medichine and Lifestyle related Disease Research Center, Kyoto University
4 Honda Research Institute Japan Co., Ltd.

3 シミュレーション実験

vM-B DNN が位相情報を直接入力して学習できることを示すため, シミュレーションによって実験を行った. 比較対象として, vM-B DNN の入力層を sigmoid 関数活性化関数とする全結合層で置き換えた B-B DNN を用意した.

3.1 データセット

Fig.1 のように仮想環境上にマイクロホンアレイと単一固定音源を設置する. マイクロホンアレイは, 半径 10cm の円周上に等間隔に 8 個のマイクロホンを配置したものとする. 音源から音を出力し, 各マイクロホンによって録音する. 得られた信号から, 周波数ごとの位相を算出し, これをデータ 1 セットとする. 正解ラベルには音源の方向を 5° 刻みに与える. これを音源の配置を変更して行い, データセットを生成する. 出力した波は, (7) 式に従って生成した. ただし, $A_i \in [0,1]$, $f_i \in (0,2000]$ はともにランダムに生成した数である. その他の詳細な実験パラメータは, Table 1 に示す. また, 学習に使用した DNN の構成は Fig.2 の通りである. ただし, vM-B Layer は, (5)(6) 式で定義したものである.

$$y = \sum_{i=1}^{100} A_i \sin(2\pi f_i t) \tag{7}$$

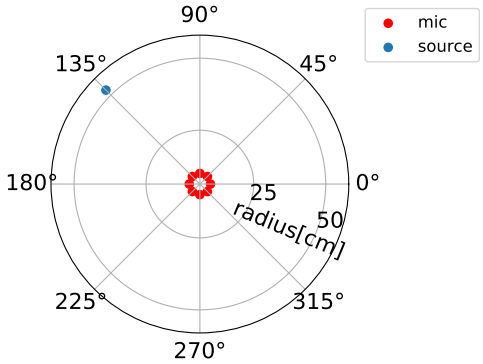

Fig. 1 マイクロホンアレイと音源の配置例 (音源配置はランダムに変更)

Table 1 実験条件

再生時間	1 秒間
周波数帯域	0〜2000Hz(256 分割)
音源距離	0.5〜1.5m
ノイズ (SNR)	20dB
学習用データ	10 万セット
テスト用データ	1 万セット

3.2 学習結果

学習結果は Fig.3 のようになった. ここで, 横軸は epoch(学習回数), 縦軸は accuracy(正解率) である. また, accuracy は, 学習モデルにおける推定値が最大となる方向と正解ラベルとの完全一致率として定義している. Fig.3 の結果から, vM-B DNN では, 通常の DNN モデルでは困難であった, 周期情報である位相情報の学習が可能であり, 音源定位ができていることが確認できる.

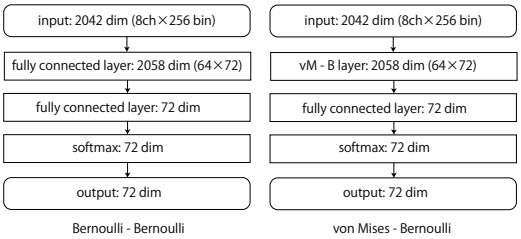

Fig. 2 DNN 構成

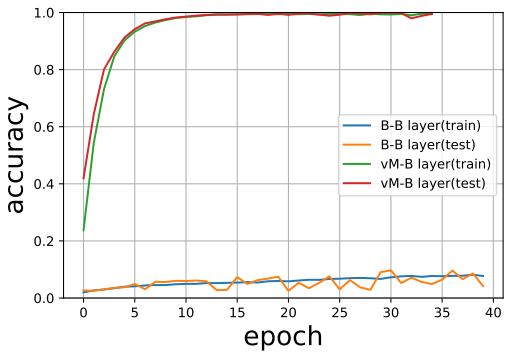

Fig. 3 実験結果 定位正解率

4 まとめ

本研究では, 音源定位に必要な位相情報を深層学習の入力として扱うため, vM-B DNN を導入した. 比較実験の結果から, 通常の DNN モデルでは学習が困難であった位相情報を学習できていることが確認できた. 今後の展開としては, 実環境における提案手法の実証実験などが挙げられる.

謝辞 本研究は, JSPS 科研費 16H02884, 16K00294, 17K00365 および, JST ImPACT タフロボティクスチャレンジの助成をうけた.

参考文献

[1] R. Schmidt, “Multiple emitter location and signal parameter estimation”, IEEE Trans. Antennas Propag. AP-34, no.3, pp.276-280, 1982.
[2] K. Nakamura, “Real-time super-resolution Sound Source Localization for robots”, IEEE, pp. 694-699, 2012.
[3] G. Hinton, “Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups”, IEEE, pp. 82-97, 2012.
[4] R. Takeda, K. Komatani, “Sound Source Localization Based on Deep Neural Networks with Directional Activate Function Exploiting Phase Information”, ICASSP, pp. 405-409, 2016.
[5] N. Yalta, K. Nakadai, T. Ogata, “Sound Source Localization Using Deep Neural Learning Models”, J. Robot. Mechatron., Vol. 29, No. 1, pp. 37-48, 2017.
[6] M. Welling, M. Rosen-Zvi, G. E. Hinton, “Exponential Family Harmoniums with an Application to Information Retrieval”, NIPS, pp. 1481-1488, 2004.
[7] K. V. Mardia, P. E. Jupp, Directional Statistics, Wiley, 1999.