

音声想起時の脳活動識別のためのデータ拡張手法の検討*

©矢野 肇, 高島遼一, 滝口哲也 (神戸大), 中川誠司 (千葉大)

1 はじめに

ブレイン・コンピュータ・インターフェイス (brain computer interface: BCI) は発話や身振りが困難な身体障害者のための意思伝達手段として研究開発が盛んに行われている。近年では、自由度の高いコミュニケーションが可能な、頭の中で想起した音声を脳活動から読み取る BCI の研究が注目されている [1]。

我々はこれまで、音声想起時の脳磁図 (magnetoencephalography: MEG) を調査し、機械学習を用いて想起音声の識別を行ってきた [2, 3, 4, 5]。先行研究 [3, 4] では、複雑な識別関数を学習可能な畳み込みニューラルネットワークが用いられたが、パラメータ数は古典的な機械学習手法よりも圧倒的に多く、実用に耐えうる高い精度を持つ識別器を個人ごとの少ない MEG データを用いて安定して訓練することは容易ではない。前稿 [5] では、複数の被験者の MEG データで訓練した条件付き変分自己符号化器を用いてデータ拡張を行い、分類精度を向上させたが、MEG の変動を再現した多様なデータを生成できたとは言い難い。

近年では、脳波 (electroencephalography: EEG) データに対してもニューラルネットワークが多く用いられるようになり、画像やテキストと比べて非常にデータ数が少なく、ノイズの多い EEG データでも安定してネットワークを訓練するため、様々なデータ拡張手法が提案されている [6, 7]。しかしながら、それらの手法の対象となっている EEG データの多くは、運動想起や睡眠など比較的多く研究されているタスクに関するものである。一般的に、タスクごとに有効なデータ拡張手法は異なり [7]、本研究が対象としている音声想起タスク、さらには MEG データに対して有効なデータ拡張手法は未だに明らかではない。

本研究では、音声想起時の脳活動データの分類精度を向上させるデータ拡張手法を調査した。音声想起時の MEG および EEG が収録された 2 つのデータセットに対し、時間・周波数・空間領域で作用するデータ拡張手法を適用し、想起された単語音声の識別をする軽量なニューラルネットワークモデルを被験者ごとに訓練した。データ拡張を用いずに訓練したネットワークとデータ拡張を用いて訓練したネットワークの分類精度を比較して、音声想起時の MEG および EEG の分類に有効なデータ拡張手法を検討した。

2 データセット

2.1 MEG データ

先行研究 [2] で収録された音声想起時の MEG データを用いた。このデータは 8 名の被験者 (男性 7 名, 女性 1 名, 20–40 歳) の MEG データからなる。被験者は、3 種類の日本語単語 (“あまぐも”, “いべんと”, “うらない”) のうちの 1 つを 2 回聴取した後、音声を聞いた通りに想起する試行を繰り返す。音声の持続時間は 800 ms で、呈示間隔は 500 ms であった。この試行は単語ごとに少なくとも 100 回行われた。試行中の MEG は被験者の頭を覆うように配置された 122 個のセンサで計測された。

計測された MEG の前処理として、サンプリング周波数が 200 Hz となるようにダウンサンプリングし、1 Hz 以下の低周波成分、1000 fT/cm を超える変動、及び眼球運動に由来するアーティファクトを除去した。頻繁に異常な信号が観測された MEG センサの信号は取り除き、他の正常なセンサの信号を元に補間した。1 回目の音声呈示及び想起のタイミングを基準に $-100-900$ ms の信号を音声聴取時及び想起時の MEG として切り出した。最後に、MEG 波形の振幅を ± 100 fT/cm が ± 1 となるように正規化した。

2.2 EEG データ

Nguyen らが公開している音声想起中の EEG データセット [1] から短い 3 種類の英単語 (“out”, “in”, “up”) を想起した 6 名の被験者の EEG データを用いた。このデータの収録では、被験者は視覚呈示された単語を 1 秒ごとに 5 回呈示されるピープ音のタイミングで想起し、ピープ音が停止した後も視覚呈示が止まるまで同じ間隔で単語の想起を続けた。この試行は単語ごとに 100 回行われた。EEG は 64 または 80 チャンネルで計測され、8–70 Hz の帯域通過フィルタ、60 Hz のノッチフィルタ、眼電図のアーティファクト除去を行った後、サンプリング周波数を 1 kHz から 256 Hz にダウンサンプリングされた。本研究では眼電図用の 4 チャンネルを除いた 60 または 76 チャンネルを用いた。各試行の最後のピープ音の呈示から 1 秒ごとに 2 秒間の EEG を合計 3 エポック取り出し、それぞれを分類器に入力するデータサンプルとした。最後に、EEG の振幅を $\pm 100 \mu V$ が ± 1 となるように正規化した。

*Data augmentation for imagined speech classification. by YANO, Hajime, TAKASHIMA, Ryōichi, TAKIGUCHI Tetsuya (Kobe Univ.) and NAKAGAWA, Seiji (Chiba Univ.).

3 データ拡張

本研究では、音声想起時の MEG および EEG データに対し、時間・周波数・チャンネル領域でデータ拡張手法を適用した。用いた手法は、睡眠時と運動想起時の EEG データセットに対して様々なデータ拡張の有効性を調べた Rommel らの研究 [7] を参考にした。また、加算平均を用いたデータ拡張手法も検討した。各領域でのデータ拡張手法について以下に述べる。

3.1 時間領域

3.1.1 符号反転 (sign flip) [8]

確率 p で全てのチャンネルの信号の符号を反転させる。脳内の活動源は電流のベクトル場で表され、観測信号の符号を全チャンネルで反転することは、全ての電流の向きを反転させることに相当する。このデータ拡張は活動源の向きが重要でない場合に有効である。

3.1.2 時間軸反転 (time reverse) [8]

確率 p で全てのチャンネルの信号の時間軸を反転させる。信号のフーリエ振幅スペクトルは信号の時間軸反転に対して不変なため、周波数成分のパワーに基づく分類では有効である可能性がある。

3.2 ガウスノイズ付加 (Gaussian noise) [9]

各時間サンプルが独立に平均 0、分散 σ^2 の正規分布に従うノイズを全てのチャンネルの信号に加算する。このデータ拡張によって脳信号に含まれるノイズに対して頑健なモデルの学習が期待できる。

3.3 加算平均 (averaging)

元の訓練データ内のクラスラベル k が付与されたデータ数を N_k とする。 N_k 個の訓練データからランダムに n 個ずつ取り出して平均した波形を新たな訓練データとする。この操作を新たな訓練データが元のデータ数の s 倍になるまで繰り返す。なお、モデルの訓練時には元の訓練データは用いない。

聴覚誘発反応や事象関連電位の解析では、外部刺激などのイベントに同期させて計測した脳信号を平均し、位相の揃っていないノイズ成分を打ち消して、目的の信号成分を得ることがしばしば行われる。このデータ拡張によって、モデルがタスクに関連した信号成分を学習し、ノイズへの過学習を防ぐことが期待できる。また、一般的に N_k から n 個のデータの取り出す組合せの数は膨大であり、多様な訓練データを生成することができる。

3.4 可変加算平均 (variable averaging)

加算平均に用いるデータ数 n を段階的に増加させながら、平均した波形を元の訓練データに追加する。具

体的には、最初は $n = 2$ とし、 N_k 個の訓練データからランダムに n 個ずつ取り出して平均していき、合計で $\lfloor N_k/n \rfloor$ 個のデータを追加する。その後、 $n \leftarrow n+1$ とし、同様の操作を $n = n_{\max}$ まで繰り返す。

上述の加算平均のデータ拡張とは異なり、訓練時に元のデータを用いるため、モデルの推論が評価データに含まれるノイズの影響を受けにくい可能性がある。また、被験者の状態の変化等でデータの信号対雑音比が変動しても、 n を変化させて様々な信号対雑音比のデータを生成することで対応できる可能性がある。

3.5 周波数領域

3.5.1 周波数シフト (frequency shift) [8]

全てのチャンネルの信号のスペクトルを一様分布 $\mathcal{U}(-\Delta f_{\max}, \Delta f_{\max})$ からサンプリングしたランダムな周波数 Δf だけシフトする。各チャンネルの信号を $x(t)$ 、その解析信号を $x_a(t) = x(t) + j\mathcal{H}[x](t)$ とする。ここで $j = \sqrt{-1}$ で、 $\mathcal{H}[x](t)$ は $x(t)$ のヒルベルト変換である。 $x(t)$ を周波数シフトした信号 $x_{\text{freqshift}}(t)$ は次のように計算される。

$$x_{\text{freqshift}}(t) = \text{Re} [x_a(t) \cdot e^{j2\pi\Delta f \cdot t}] \quad (1)$$

ここで、 $\text{Re}[\cdot]$ は複素数の実部を取り出す演算である。

このデータ拡張は、一部の脳信号が持つ特定の周波数のピークが変動する場合に有効である。

3.5.2 Fourier transform surrogate [10]

フーリエ位相スペクトルを周波数ごとに一様分布 $\mathcal{U}(0, \Delta\phi_{\max})$ からサンプリングしたランダムな位相 $\Delta\phi$ だけシフトする。各周波数における $\Delta\phi$ は全チャンネルで共通である。各チャンネルの信号 $x(t)$ のフーリエ変換を $\mathcal{F}[x](f)$ とすると、位相をランダム化した信号 $x_{\text{fts surrogate}}(t)$ のフーリエ変換は次の式で表される。

$$\mathcal{F}[x_{\text{fts surrogate}}](f) = \mathcal{F}[x](f)e^{j\Delta\phi} \quad (2)$$

処理後の時間信号 $x_{\text{fts surrogate}}(t)$ は式 (2) の逆フーリエ変換によって得られる。

このデータ拡張によって元の信号と同じ振幅スペクトルを持つ異なる波形の信号を作ることができるため、モデルに信号の時間波形よりも周波数成分を重視するように訓練させることができると考えられる。

3.6 空間領域

3.6.1 チャンネル左右反転 (channel symmetry) [11]

確率 p で頭の左半球および右半球に配置された全てのチャンネルの信号を頭の正中線に関して対称な位置にあるチャンネルの信号に入れ替える。分類したいタスクが特定の半球のみに出現する脳活動に関連していない場合、この手法は有効であると考えられる。

3.6.2 Channel dropout [12]

チャンネルごとに確率 p で信号を 0 にする。モデルによる予測が特定のチャンネルの信号に依存しすぎないようにすることで、モデルの過学習を抑え、汎化性能を向上できると考えられる。

4 評価実験

4.1 実験条件

音声想起時の MEG および EEG の分類モデルとして、EEGNet [13] を被験者ごとに訓練した。EEGNet は一般的な畳み込みの代わりに depthwise separable convolution を用いた軽量なネットワークであり、データ数が少量でも訓練が可能である。EEGNet は時間方向の畳み込み層、センサ方向の depthwise convolution 層、時間方向の separable convolution 層、及び全結合層からなる。本研究では、EEGNet の構造を変更し、separable convolution 層を 2 層にして用いた。

訓練したモデルを 10 分割交差検証によって評価した。評価データセット以外のデータセットの 80% を訓練データセットに、20% を検証データセットに分割した。データセットの分割時、MEG データセットでは各分割内の試行が時間的に連続するように分割し、EEG データセットでは各分割内で 3 つのクラスラベルの数がそれぞれ等しくなるようにランダムに分割した。学習されたモデルの評価指標には macro F1 score を用いた。この指標はクラスごとに算出した F1 score を平均したものである。EEGNet は最大で 200 エポック訓練し、検証データセットに対して macro F1 score が最大となるモデルを選択した。

加算平均と可変加算平均のデータ拡張は MEG データセットのみに適用した。これらの手法は予め訓練データセット全体に適用され、適用後のデータを集めた新たな訓練データセットがモデルの訓練に用いられた。その他のデータ拡張手法はデータをモデルに入力する直前に適用された。Channel dropout によるデータ拡張は各チャンネルの位置が与えられている MEG データセットのみに適用した。

4.2 結果

音声想起時の MEG データにデータ拡張を行って訓練したモデルの macro F1 score の、データ拡張を行わなかった場合からの改善量を Table 1 に示す。MEG データに対して多くの被験者で macro F1 score が向上したデータ拡張手法は、時間軸逆転、ガウスノイズ付加、加算平均、channel dropout であった。一方で、符号反転、周波数シフトと Fourier transform surrogate は、データ拡張を用いなくとも macro F1 score が高かった被験者 4, 5, 6, 8 のモデルの macro

F1 score を大きく低下させた。

音声想起時の EEG データにデータ拡張を行って訓練したモデルの macro F1 score の、データ拡張を行わなかった場合からの改善量を Table 2 に示す。EEG データに対して平均的にモデルの macro F1 score を向上させたデータ拡張手法は、符号反転と Fourier transform surrogate であった。時間軸逆転は被験者 8 のモデルの macro F1 score を大きく低下させたものの、被験者 3, 5, 6, 12 のモデルの macro F1 score を向上させた。これらのデータ拡張手法を組み合わせた場合、被験者 8 以外の被験者のモデルの macro F1 score が向上した。一方で、周波数シフトや channel dropout を用いた場合、多くの被験者のモデルで macro F1 score が低下した。

4.3 考察

各データ拡張手法の効果はデータセットおよび被験者の間で一貫していなかった。データセット間での違いは主に MEG と EEG の違いによるものと考えられる。脳内の活動源で発生した電気信号は、脳内の組織ごとの電気伝導率の違いから電極で EEG として観測されるまでに拡散し歪む一方で、活動源における電流が生じさせる磁気的な信号は、組織ごとの透磁率が真空の透磁率とほぼ等しいため、ほとんど歪まずに磁気センサで MEG として観測される。このため、MEG と EEG で適切なデータ拡張手法が異なる可能性がある。被験者間での違いは音声想起時の脳活動の個人差によるものと考えられる。特に時間軸逆転は、いずれのデータセットでも分類精度が向上する被験者と低下する被験者が観察された。性能が向上した被験者では音声想起時の脳信号に左右対称のパターンが出現していた可能性がある。

2 つのデータセットに対する結果から、分類精度の高かった被験者のモデルは、音声想起時に出現した脳信号の時間波形パターンを学習した可能性が示唆される。まず、符号反転はいずれのデータセットでもデータ拡張を用いない場合に分類精度が比較的高かった被験者のモデルの精度を低下させた。このことは、モデルが音声想起時に出現した特定の向きを持った活動源の波形を学習したことを示唆しており、特に歪みが少ない MEG データに対する結果にその傾向が現れたと考えられる。次に、MEG データに対して加算平均のデータ拡張を適用した場合にモデルの分類精度が向上したが、可変加算平均を適用した場合にはあまり精度向上は見られなかった。いずれの手法も訓練データ数は元の約 3 倍であるが、加算平均を適用した訓練データの方が、試行間で位相同期した成分が強調されていたと考えられる。さらに、MEG データに対して周波数成分ごとの位相のランダム化で分

Table 1 Amount of the macro F1 score improvement for the MEG dataset

| Subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Mean |
|---|------|------|------|-------|-------|------|------|------|------|
| w/o augmentation | 32.5 | 29.6 | 31.0 | 41.1 | 39.9 | 38.3 | 29.4 | 36.8 | 34.8 |
| (1) Sign flip ($p = 0.5$) | -2.4 | +3.5 | +3.7 | -11.3 | -7.2 | -4.6 | -2.2 | -4.0 | -3.1 |
| (2) Time rev. ($p = 0.5$) | +2.2 | +4.0 | -2.8 | -3.1 | +0.0 | -4.5 | +2.7 | +5.6 | +0.5 |
| (3) Gauss noise ($\sigma = 0.2$) | +3.8 | -0.3 | +0.7 | -2.5 | +4.0 | -1.9 | +2.0 | +4.7 | +1.3 |
| (4) Ave. ($n = 10, s = 3$) | +3.8 | +1.5 | -1.0 | -0.4 | +4.0 | -0.9 | +6.2 | +4.7 | +2.2 |
| (5) Var. ave. ($n_{\max} = 10$) | +2.1 | -0.4 | -1.5 | -0.7 | -0.8 | -2.4 | +4.1 | +3.1 | +0.4 |
| (6) Freq. shift ($f_{\max} = 0.5$) | +2.8 | +2.2 | +3.4 | -10.5 | -10.7 | -6.7 | +4.3 | -6.8 | -2.7 |
| (7) FT surrogate ($\Delta\phi_{\max} = 2\pi$) | +2.8 | +4.5 | -2.5 | -12.2 | -9.0 | -6.4 | +0.0 | -7.8 | -3.8 |
| (8) Ch. sym. ($p = 0.5$) | +2.5 | -2.3 | -1.0 | -0.3 | -3.9 | -0.0 | +5.0 | +4.1 | +0.5 |
| (9) Ch. drop. ($p = 0.4$) | +1.0 | +3.3 | -0.4 | +2.6 | +0.6 | +0.1 | +1.7 | +6.9 | +2.0 |
| (10) (2)+(3)+(4)+(8)+(9) | -1.8 | +4.1 | -0.4 | -6.3 | -5.2 | -1.6 | +7.4 | +3.5 | -0.1 |

Table 2 Amount of the macro F1 score improvement for the EEG dataset

| Subject | 1 | 3 | 5 | 6 | 8 | 12 | Mean |
|---|------|------|------|------|------|------|------|
| w/o augmentation | 35.6 | 33.4 | 34.1 | 34.9 | 32.4 | 44.1 | 35.7 |
| (1) Sign flip ($p = 0.5$) | -1.2 | +2.4 | +4.4 | -1.9 | +2.6 | -3.2 | +0.5 |
| (2) Time rev. ($p = 0.5$) | -1.4 | +3.6 | +2.2 | +1.1 | -8.9 | +0.1 | -0.6 |
| (3) Gauss noise ($\sigma = 0.4$) | -2.5 | +0.0 | +0.8 | -3.2 | +0.4 | -5.3 | -1.6 |
| (4) Freq. shift ($f_{\max} = 0.5$) | -0.4 | +4.3 | +1.5 | -0.2 | -3.8 | -9.6 | -1.4 |
| (5) FT surrogate ($\Delta\phi_{\max} = 2\pi$) | +2.6 | +3.1 | +4.8 | -0.9 | -5.1 | +1.1 | +0.9 |
| (6) Ch. drop. ($p = 0.4$) | -1.1 | -1.9 | +1.8 | -0.5 | -0.2 | -4.5 | -1.1 |
| (7) (1)+(2)+(5) | +1.9 | +4.6 | +3.6 | +1.5 | -0.7 | +1.1 | +2.0 |

類精度が低下したことは、音声想起時の脳信号の位相スペクトルが分類に有効であることを示している。一方で、EEG データに対しては位相をランダムにした場合に精度の向上が見られたが、EEG の方が伝播の過程で位相が歪んでおり、データ拡張による位相スペクトルへの影響が小さかったためと考えられる。本研究で用いた周波数シフトのデータ拡張も音声想起時に出現する脳信号の時間波形を歪めてしまったために分類精度の低下を引き起こした可能性がある。

Channel dropout は MEG データセットに対してはモデルの分類精度を向上させたが、EEG データセットに対しては分類精度を低下させた。EEG データよりも MEG データの方がより多くのチャンネルで計測されたものであり、伝播時に波形が歪まないため、0 にされたチャンネルの情報を他のチャンネルから補いやすかったと考えられる。このため、MEG データで訓練されたモデルは channel dropout によって過学習が防がれ、性能が向上したと考えられる。

5 おわりに

本研究では、音声想起時の MEG および EEG が収録された 2 つのデータセットに対し、時間・周波数・空間領域で作用するデータ拡張手法を適用し、想起音声の分類に有効なデータ拡張手法を調査した。その結果、MEG データでは加算平均と channel dropout を用いたデータ拡張が、EEG データでは Fourier transform surrogate を用いたデータ拡張が分類精度の向上に寄

与した。また、音声想起時の脳信号の位相を保つデータ拡張手法が想起音声の分類に有効である可能性が示唆された。

謝辞 本研究の一部は、JSPS 科研費 JP22K18626 の支援を受けて実施された。

参考文献

- [1] C. H. Nguyen *et al.*, J. Neural Eng., 016002, 2018.
- [2] S. Uzawa *et al.*, IEEE EMBC, 2542–2545, 2017.
- [3] 矢野ら, 音講論 (春), 507–510, 2020.
- [4] 山名ら, 音講論 (春), 517–520, 2023.
- [5] 矢野ら, 音講論 (春), 645–648, 2024.
- [6] C. He *et al.*, Front. Hum. Neurosci., 15, 765525, 2021.
- [7] C. Rommel *et al.*, J. Neural Eng., 066020, 2022.
- [8] C. Rommel *et al.*, ICLR, 2022.
- [9] F. Wang *et al.*, MultiMedia Modeling, 82–93, 2018.
- [10] J. T. C. Schwabedal, *et al.*, arXiv:1806.08675, 2019.
- [11] O. Deiss *et al.*, arXiv:1803.09702, 2018.
- [12] A. Saeed *et al.*, IEEE ICASSP, 1255–1259, 2021.
- [13] V. J. Lawhern *et al.*, J. Neural Eng., 15, 056013, 2018.