

# 地震予知を目的とした DEMETER 衛星データの雷ノイズ除去

石通彩加<sup>1</sup> 竹本有紀<sup>1</sup> 石川由羽<sup>2</sup> 高田雅美<sup>1</sup> 城和貴<sup>1</sup>

**概要:** 本稿では、ディープクラスタリングによる DEMETER 衛星データから雷ノイズを除去する手法を提案する。地震大国である日本において、被害を最小限に抑えるためにも地震短期直前予測の早期実現が望まれている。電磁波電場強度を記録した DEMETER 衛星データに工学的手法を適用することで、地震予測が可能になると考えられる。しかし、データ内には雷が強度なノイズとして混在しているため、まずはノイズ除去を行う必要がある。そこで、ディープクラスタリングを用いてノイズとして入り込んだ雷を除去する手法を提案する。混合データである衛星データから、雷ノイズと電磁波電場データに分離する実験を行い、提案手法の有用性を確認する。

**キーワード:** ディープクラスタリング, ノイズ除去, 地震短期直前予測, 短時間フーリエ変換, BLSTM

## For the purpose of earthquake prediction Lightning noise removal from DEMETER satellite data

AYAKA ISHIDORI<sup>†1</sup> YUKI TAKEMOTO<sup>†1</sup>  
YU ISHIKAWA<sup>†2</sup> MASAMI TAKATA<sup>†1</sup>  
KAZUKI JOE<sup>†1</sup>

### 1. はじめに

日本は世界でも珍しく、4枚のプレート上に位置しており、世界で発生する地震の約2割は日本付近で発生している[1]。そのため、日本は地震大国と呼ばれている。実際に2018年に日本付近で発生した地震は2179回にのぼり、最大震度4以上を観測した地震は78回起きている。2011年に起きた東日本大震災で、多大な被害が出たことは人々の記憶に新しい。地震の発生を事前に察知することができれば、地震被害を減らすことができる。地震予測に関する人々の期待は高まっている。地震予測の方法には、統計的手法を用いる方法や、地震前に発生する先行異常現象を用いて予測する方法がある[2]。これらの地震予測は予測時間によって分類することができる。有用な予測方法として、長期予測と直前予測が広く認められている。長期予測は、数年から数十年後に地震が発生する可能性を予測する。政府の地震調査研究推進本部がプレートの境界や断層の位置、地震の歴史データに基づき、地震の発生確率を公表している。現在、南海トラフ地震が30年以内に発生する確率は70%以上であることが判明している。直前予測は、地震の発生直後に震源付近の観測点にある地震計で捉えた初期微動を解析することで、主要動の到達時間や地震の規模を知らせる予測である。そのため、地震を感じる数秒から数十

秒前に予測が行われる。これは緊急地震速報として既に実用化されている。これら2つの予測情報は予測時間が極端であり、地震に対して十分に備えることが難しい。東日本大震災以降、人々の防災に対する意識は高まっており、地震による被害を最小限に食い止めるためにも、数週間から数日前に地震が発生することを知らず地震短期直前予測システムの早期実現が望まれている。

地震に先行して、様々な物理化学的異常現象が起こると考えられている。その中に、電磁気学的な現象が存在し、地震の直前予測に有望である。電磁気学的な現象は、地震前に発生する電気や電波、磁気の異常を測定して分析される。本稿では、予測に夜間 VLF 帯電磁波電場強度現象[3][4]という先行現象を用いる。地上から約90kmの高さにある電離圏に VLF 帯電磁波が入射すると、震央付近上空の電場強度が減少する。この先行現象から、地震発生から発生前4時間以内に1.7kHz帯周辺の電場強度が4~6dB減ることが統計的に明らかとなっている。これは、電場強度を記録した DEMETER 衛星データに地震前駆シグナルとして表れる。衛星データに、ディープラーニングを始めとする工学的手法を適用すると、地震前駆シグナルが抽出でき、地震の短期直前予測が可能になると考えられている。しかし、データ内には雷が強度なノイズとして混在するため、まずはノイズ除去を行う必要がある。本稿では、DEMETER 衛星データのノイズ除去としてディープクラスタリングを用いた手法を提案し、実験を行う。

本稿の構成は以下の通りである。2章では既存研究とし

<sup>1</sup> 奈良女子大学  
Nara Women's University  
<sup>2</sup> 滋賀大学  
Shiga University

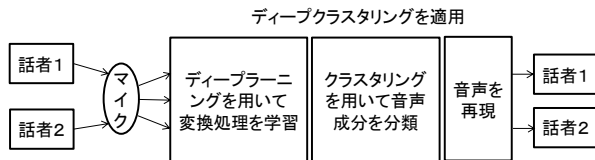


図1：ディープクラスタリングを用いた音声分離の流れ

て、ディープクラスタリングを用いた1つのマイクで記録した混合音声の分離について述べる。3章では、DEMETER衛星データにディープクラスタリングを適用したノイズ除去手法の提案を行い、4章では、実験方法と結果を記す。

## 2. 既存研究

三菱電機株式会社は、ディープクラスタリングを用いて単一の信号源で記録される混合音声から各音声を分離し、再現する研究を行っている[5][6]。雑音などのノイズが混在した音声データからノイズ除去を行う技術は、既にカーナビゲーションなどに搭載されているハンズフリー通話機能で実用化されている。複数話者が同時に話す声を聞き分けて再現する技術は2つあげられる。1つ目は、複数マイクを用いて話者の位置情報を取得することで音声を分離、再現する技術である。もう1つは、マイク1本で録音した音声を判別して分離し、再現する技術である。後者は再現率が低く、技術力の向上が期待されている。既存研究[5][6]では、マイク1本で録音した2者および3者が同時に話す音声を分離し、再現している。2者の同時音声では従来の再現率が51%であるのに対し、原音再現率90%以上を達成している。音声は事前に登録されていないものであり、女性同士や男性同士だけでなく、異なる言語の同時音声にも対応している。音声分離に使用されるディープクラスタリングは、教師あり学習のディープラーニングと教師なし学習のクラスタリングを組み合わせた技術である。図1に、既存研究で適用されたディープクラスタリングの流れを示す。初めに、ディープラーニングを用いて音声成分の持つ特徴から話者を分類する変換処理を学習する。次に、学習した変換処理を入力音声に適用する。その後、k-meansクラスタリングを行うことで音声成分を分離する。最後に、分離した音声成分を再合成することで各話者の声を再現する。音声成分の学習は、抑揚や音の流れといった音声特徴の類似性をもとに行う。

本稿は、単一の観測点から取得される時系列データを対象とするため、既存研究のディープクラスタリングを用いたノイズ除去手法を提案する。

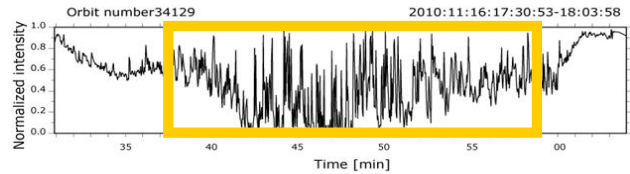


図2：DEMETER衛星1軌道中に混在する雷ノイズ

## 3. 提案手法

### 3.1 夜間 VLF 帯電磁波電場減少

本稿では地震短期直前予測の先行現象として、夜間 VLF 帯電磁波電場減少を使用する。現在、地震が発生する前に電離圏が数 km 低下することが判明している。電離圏は地上から約 90km の高さであり、高度によって電子密度が異なる。太陽からの紫外線を受けて酸素や窒素は電離し、高度が高くなるにつれて電離量が増える。本稿で扱う先行現象は、D 領域と言われる電離圏下部で行われる。D 領域は電離圏の中では最も大気が濃く、電離量も少ない。ここに電磁波が入射すると、電磁波周りの電場が電離圏の電子を揺さぶる。しかし、大気が濃いためにジュール熱で電場強度が減少する特徴がある。DEMETER 衛星データが記録した電磁波電場強度と衛星軌道近傍に発生した地震との関係を調べた結果、以下のことが統計的に明らかとなっている。地震が発生する4時間前以内に周波数 1.7kHz 帯周辺において、電磁波電場強度が 4~6dB ほど減少する。また、乱数で仮想的に発生させた仮想地震の場合は電場強度の変化が地震発生前に確認されない。この特徴を夜間 VLF 帯電磁波電場強度減少と呼び、地震の先行現象として用いる。本稿では、この強度変化を DEMETER 衛星データ内に表れる地震前駆シグナルとする。

### 3.2 DEMETER 衛星データの問題点

衛星データを採取する DEMETER は高度 660km の太陽同期準回帰軌道上で運用されていた衛星である。本稿では、VLF 帯 Burst mode のデータを用いる。Burst mode では地震活動が活発な領域に対してデータを取得する。水平電場 1 成分、磁場 1 成分についてサンプリング周波数 40kHz で観測する波形データである。DEMETER が運用期間に採取したデータ総量は 1.3TB にのぼる。図2は、DEMETER 衛星が1軌道を動く間に測定した 1.7kHz 帯のデータを規格化したものである。縦軸は電場量の強度変化を表し、横軸は時間を示す。黄色枠の中に振幅の大きい値が見られる。本稿では地球の周りを南から北方向に移動する軌道を用いる。DEMETER 衛星が移動する際に、極を離れるにつれて雷が多く観測される。データ内に見られる値の大きな振幅は雷がノイズとして混在したものである。地震前駆シグナルは、夜間に記録する DEMETER 衛星データ内に雷ノイズと共に含まれる。それに加えて、雷の強度変化の方が地震前駆シ

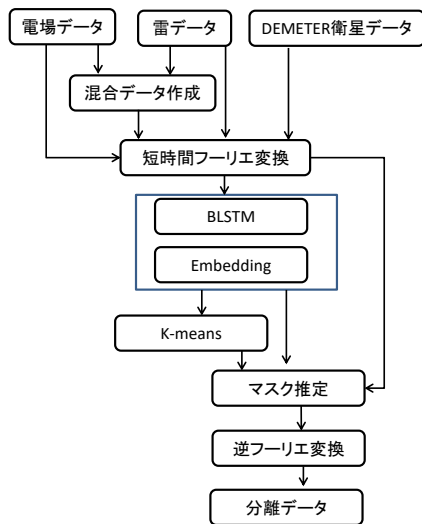


図3：提案手法の流れ

グナルで表れる変化量よりも激しい。こうした問題点から、このままでは大量にある観測データを活用することができない。データから地震前駆シグナルを抽出して地震予測を行うために、まずは観測データに混在する雷ノイズを除去する必要がある。そこで、本稿は DEMETER 衛星データ内にノイズとして混入する雷をディープクラスタリングで除去する手法を提案する。

### 3.3 提案手法の流れ

DEMETER 衛星の観測データは、電場量に加えて雷ノイズが混入している。また、巨大地震発生時には地震前駆シグナルも含まれる。観測データに既存のノイズ除去手法を適用すると、雷ノイズだけでなく地震前駆シグナルまで除去してしまう可能性がある。提案手法は、地震前駆シグナルを含まない電場データを雷ノイズとノイズ以外に分ける。地震前駆シグナルを含む電場データに対しても同様に提案手法を施すと雷ノイズとノイズ以外の部分に分離できる。この時、地震前駆シグナルはノイズ以外の部分に表れると考える。提案手法のモデルを図3に示す。また、提案手法は以下の手順に従って行う。

- (1) 地震前駆シグナルの有無による軌道選択
- (2) データ作成
  - (a) DEMETER 衛星データ作成
  - (b) 電場データ作成
  - (c) 雷データ作成
  - (d) 混合データ作成
- (3) 短時間フーリエ変換を用いた周波数成分の抽出
- (4) 周波数特徴の学習
- (5) クラスタリングによるデータ生成

手順 (1) では、地震前駆シグナルの有無によって軌道の選択を行う。手順 (2) では、手順 (1) で選んだ軌道のデー

タから 4 種類のデータを作成する。DEMETER 衛星データ、電場データ、雷データ、そして電場データと雷データからランダムに合成した混合データである。電場データ、雷データ、混合データは手順 (3) で求める周波数成分を学習に用いるデータセットとする。データセットは教師データと検証データに分け、教師データを用いて学習後、検証データによって学習の精度を確認する。DEMETER 衛星データは、分離の精度評価にテストデータとして使用する。手順 (3) では、短時間フーリエ変換を用いて、電場データ、雷データ、混合データの周波数成分の特徴量としてスペクトログラムを計算する。また、精度評価では DEMETER 衛星データのスペクトログラムを計算する。手順 (4) では、ディープラーニングを用いて、周波数特徴の埋め込みベクトル空間と理想とするバイナリマスクとの KL 情報量を最小化するように学習を行う。手順 (5) では、学習から学んだ変換処理をテストデータに適用し、データ間の類似度によってクラスタリングする。最後に、逆フーリエ変換を用いて時間周波数ビンごとに、分離後のデータを生成する。手順の詳細は 3.4 節から 3.8 節で述べる。

### 3.4 地震前駆シグナルの有無による軌道の選択

提案手法に使用する、電磁波電場強度を記録した DEMETER 衛星の観測データは、電場量に加えて雷ノイズが混入している。また、巨大地震発生時には地震前駆シグナルもデータ内に含まれる。従来の手法で電場データに対してノイズ除去を行おうとすると、未知である地震前駆シグナルを残したまま雷ノイズを除去することができない。まずは、DEMETER 衛星データに提案手法を施し、雷ノイズのない電場データを取得する。その後、地震前駆シグナルの抽出を目指す。本稿では、地震前駆シグナルを含まないデータに提案手法を適用して、雷ノイズを正確に除去できるように学習する。そして、地震前駆シグナルを含む軌道にも提案手法を施し、データから雷ノイズのみを分離できるか確認する。提案手法に用いる地震前駆シグナルを含まないデータは、地震による影響を受けていない軌道を選択することで得られる。地震前駆シグナルを含む軌道、含まない軌道の選択方法を説明する。巨大地震発生時、地震前駆シグナルは DEMETER 衛星データ内に雷ノイズと共に含まれている。地震前駆シグナルが表れる先行現象は統計的に、巨大地震が起こるときに発生する。MySQL を用いて、M4.8 以上の地震と以下の条件を照らし合わせ、地震前駆シグナルを含む軌道を選択する。地震記録は USGS 地震カタログ[7]に従う。

- (1) 地震発生時刻 4 時間前から地震発生時刻までの間に、震源から 330km 半径の円内を DEMETER 衛星が通過している軌道
- (2) (1) を満たす軌道のうち、震源最近接時刻から前後 250 秒の時刻を確保できるもの

条件 (1), (2) を満たす軌道はデータ内に地震前駆シグナルが混入しており, 地震前駆シグナルを含む軌道と言える. そして, 地震前駆シグナルを含まない軌道は条件 (1) (2) を満たさないものである.

### 3.5 データ作成

提案手法に使用するデータの作成方法について述べる. 軌道の選択後, 提案手法の入力データとして, DEMETER 衛星データ, 電場データ, 雷データ, 混合データを作成する. DEMETER 衛星データは, 電場データに雷がノイズとして混入しているデータである. 学習結果の精度を検証するテストデータとして, DEMETER 衛星データを使用する. 電場データと雷データは, 学習に用いるデータセットとして使用する. 混合データは, ランダムに選ばれた電場データと雷データから作成される. 混合データの周波数成分もデータセットとして用いる. 各データの作成手順の詳細については 3.5.1 項から 3.5.4 項で説明する.

#### 3.5.1 DEMETER 衛星データ

DEMETER 衛星が観測したデータは Universal Time の年, 月, 日, 時, 分, 秒, ミリ秒, 軌道番号, 緯度, 経度, 磁気緯度, 磁気経度に加えてサンプリング周波数 40kHz で水平電場 1 成分と磁場 1 成分を連続的に記録している[8]. 先行現象である夜間 VLF 帯電磁波電場減少は 1.7kHz 帯で発生する. そのため, 該当する周波数のデータのみを抽出する必要がある. 周波数帯は 16 種類に分類される. 本稿では 1.7kHz 帯に分類される 1621.09375Hz~1718.75000Hz を使用する. DEMETER 衛星データの作成手順を述べる. まず初めに, 観測データに高速フーリエ変換 (Fast Fourier Transform) を適用し周波数スペクトルを算出する. 次に, 1.7kHz 帯に該当する周波数のみを通すバンドパスフィルタを用いて周波数帯の選定を行う. 最後に逆高速フーリエ変換で波形に戻す. この整形作業を行った 1.7kHz 帯の周波数帯電場データを DEMETER 衛星データとする. フーリエ変換は, サンプリング周波数 40kHz, サンプル数 256, スライド幅 50%, 窓関数をブラックマン・ハリスに設定する. この窓関数は, DEMETER 衛星データのスペクトル解析の際に用いられることから本稿でも使用する. ハニング窓やハミング窓に比べてスペクトルの漏れを抑制する特徴を持つ.

#### 3.5.2 電場データ

電場データは, 3.5.1 で作成した DEMETER 衛星データにパーセンタイルを用いて算出する. パーセンタイルは, データを昇順に並べた替えた時, 小さい方から数えて全体の  $p\%$  に位置する値を  $p$  パーセンタイル ( $0 \leq p \leq 100$ ) と呼ぶ. データ数  $n$  の  $p$  パーセンタイル位置を  $P$ ,  $p$  パーセンタイルの値を  $A$  とする.  $P$  が小数の時,  $P$  の整数部分を  $\rho$ , 小数部分を  $r$  とする.  $n$  個のデータを小さい順に並べた際に  $\rho$  番目,  $\rho + 1$  番目のデータを  $D_\rho, D_{\rho+1}$  とおくとパーセンタイルは式 1, 式 2 で与えられる.

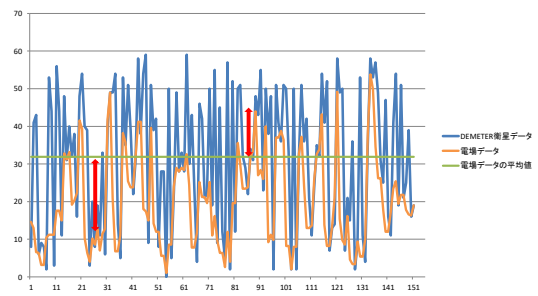


図 4 : 雷データ作成時の参考図

$$P = (n + 1)p/100 \quad (1)$$

$$A = D_\rho + r(D_{\rho+1} - D_\rho) \quad (2)$$

電場データはデータ数 40, 10 パーセンタイルの計算結果とする. DEMETER 衛星データの 1 軌道を 1 ファイルとみなす. 式 1, 式 2 を使用して算出された値を電場データとして用いる.

#### 3.5.3 雷データ

雷データを作成する手順を説明する. 図 4 は雷データを作成する際の参考図である. 青線を DEMETER 衛星データ, オレンジ線を電場データ, 黄緑線を電場データの平均値とする. まず初めに, 電場データの平均値を求める. 次に, その平均値と電場データとの差を計算する. 図 4 の赤矢印の範囲が  $\alpha$  である. DEMETER 衛星データの各値が平均値より大きい時に  $\alpha$  を減算し, 平均値よりも小さい時に  $\alpha$  を加算する. 演算後のデータを雷データとして使用する.

#### 3.5.4 混合データ

電場データ, 雷データ, 混合データ 1 個ずつの計 3 個 1 組をデータセットとする. このデータセットを入力データとして用いる. 電場データと雷データをランダムに選択し, 混合データを作成する. 混合データの長さは, 電場データと雷データの長さを比較して短いデータの長さに合わせる. 教師データの組み合わせ数と同じ数の混合データを作成する.

### 3.6 短時間フーリエ変換を用いた周波数成分の抽出

電場データ, 雷データ, 混合データに短時間フーリエ変換 (Short-Time Fourier Transform) を適用し, 周波数成分の特徴量であるスペクトログラムを計算する. 短時間フーリエ変換は, 音声の時間周波数分析に用いられている信号処理法の 1 つである. 特徴抽出, 雑音除去, 音源分離などの変換や合成を含む様々な信号処理に幅広く用いられている. まず初めに, 一定の大きさの窓関数を用いて信号を切り出す. 次に, 切り出したフレーム部分にフーリエ変換を行い, 各フレームの周波数成分を求める. そして, 求められた値からスペクトル計算を行う. その後, 次々にフレームをずらして同じ処理を繰り返す. 設定した 1 つのフレームに対して 1 組のスペクトルが得られるため, 周波数成分の時間



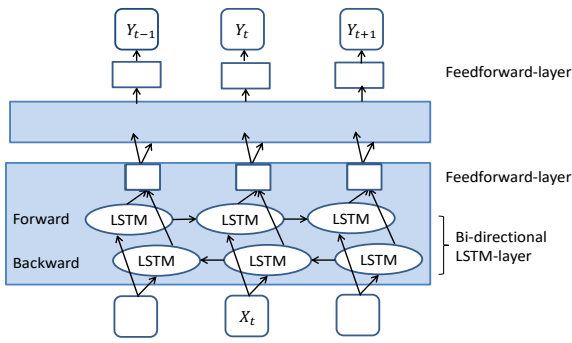


図 5: 提案手法に用いる BLSTM モデル

的变化を知ることができる。計算式は以下のように表される。

$$X(t, \omega) = \int_{-\infty}^{\infty} x(T) w(T-t) e^{-j\omega T} dT \quad (3)$$

窓関数  $w(t)$  を時刻 0 が中心となるように定義する。式 3 から得られる時間周波数スペクトル  $X(t, \omega)$  は複素関数であり、その位相は時間  $t$  に依存して変化する。短時間フーリエ変換で求めた値  $X(t, \omega)$  の絶対値を 2 乗し、対数をとった値を特徴量として用いる。短時間フーリエ変換では、時間分解能や周波数分解能はフレームの特徴に大きく依存する。本稿では、サンプリング周波数 40,000、フレームサイズは 2048 とする。窓関数にはハンギング窓を用いて、スライド幅はフレームサイズの 25% とする。

### 3.7 周波数特徴の学習

ディープラーニングの分野において、多様なネットワークが提案されている。音声波形や文章などの時系列データに対して有効なネットワークに RNN (Recurrent Neural Network) がある。本稿では、RNN の中でも、長期依存性を持つ LSTM[9] を双方向に発展させた BLSTM (Bi-directional Long Short Term Memory) [10] を用いる。図 5 にニューラルネットワークの構造を示す。2 層の BLSTM 層とフィードフォワード層で構成される。各 BLSTM 層は 600 の隠れセルとフィードフォワード層を持つ。フィードフォワード層は 40 の埋め込み次元を設定する。入力には、3.6 節で求めた特徴量が使用され、各時間周波数ビンの埋め込みベクトルを出力する。埋め込み行列  $V \in R^{TF \times D}$  は、 $T$  はフレーム数、 $F$  は周波数ビンの数、 $D$  は埋め込み次元を指す。学習は  $V$  が、理想的なバイナリマスク  $Y \in R^{TF \times C}$  に近づくように行われる。 $Y$  は正解のクラスタリング結果を 0,1 で表現し、 $C$  はデータ内の混合物数と一致する。損失関数  $L$  は式 4 に表す。アフィニティ行列  $VV^T$  と  $YY^T$  の差を計算し、学習は損失関数が最小化するように行う。

$$L = \|VV^T - YY^T\|_F^2 \quad (4)$$

また、VAD 閾値を 40dB と設定し、VAD より大きい時間周波数ビンに対応する埋め込みのみが損失計算に寄与する。パラメータの最適化は、確率的勾配降下法で行う。最適化関数には adam (Adaptive Moment Estimation) を用いて自動調整する。学習の評価には、最小二乗法を用いて算出した誤差の loss データを使用する。

### 3.8 クラスタリングによるデータ生成

学習後、k-means を用いてクラスタリングを行う。埋め込みベクトルを雷データと電場データの各クラスターに分割する。テストデータは、データを読み込んだ後に各周波数ビンがどのデータに所属しているのかを識別する。時間周波数マスクを推定し、逆短時間フーリエ変換とオーバーラップ加算処理を用いて時間領域で分離された信号をデータとして取得する。

## 4. 実験

### 4.1 実験方法

本章では、提案手法を用いて DEMETER 衛星データを雷データと電場データに分離する実験を行う。本実験の実行 OS は、Ubuntu16.04 である。使用する計算機は、Intel® Core™ i7-7700 CPU @ 3.60GHz×8 である。GPU は NVIDIA GeForce GTX 1080 Ti を用いる。機械学習のライブラリには、ディープラーニングのフレームワークである TensorFlow[11] を利用する。

本実験で使用するデータについて述べる。データは、静岡県立大学の鴨川仁准教授より提供されたものである。本実験では DEMETER 衛星が水平電場 1 成分、磁場 1 成分についてサンプリング周波数 40kHz で取得する VLF 帯 Burst mode のデータを用いる。連続的に測定した時系列データであり、2006 年 6 月から 2007 年 12 月に観測されたものを使用する。地震前駆シグナルの有無による軌道選択後に DEMETER 衛星データ、雷データ、電場データ、混合データを作成する。雷データ、電場データ、混合データ 1 つずつ、計 3 つのデータを 1 セットとする。100 セットを教師データとして使用し、10 セットを検証データとして使用する。また、テストデータとして DEMETER 衛星データ 2 つを分離結果の精度確認に用いる。学習時のバッチサイズは 64 である。

### 4.2 実験結果と考察

本稿では、DEMETER 衛星データ内に混在する雷ノイズの除去をディープクラスタリングで試みる。地震前駆シグナルを含まない場合、含む場合の結果を述べる。図 6 は地震前駆シグナルを含まない軌道の DEMETER 衛星データから雷ノイズを除去した結果である。縦軸は電場量、横軸はデータ点数を示す。(1) は学習開始直後、(2) は学習開始から 250 エポック後、(3) は正解とする電場量である。図 6 (1) より、学習開始直後は大まかな形状を捉えているが

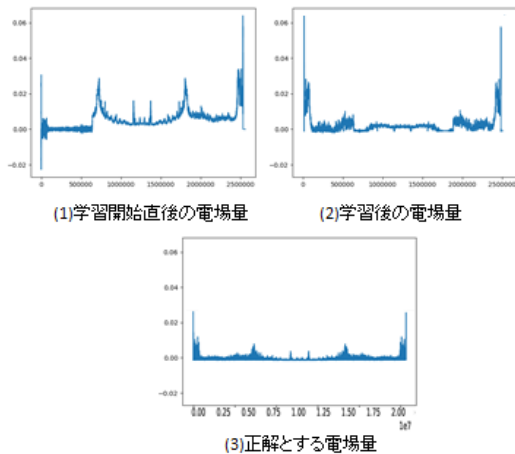


図 6：提案手法による雷ノイズ除去後の電場量

誤認識している値が多くみられる。学習が十分に行われていない事がわかる。250 エポック目では、学習直後に比べて電場の形状を捉えている。しかし、正解とする電場量に比べて出力値が大きい。また、未判定の部分やクラスタリング間違いが確認できる。次に、地震前駆シグナルを含む軌道にも提案手法を適用すると、同様に雷ノイズとそれ以外に分離している事が確認できる。地震前駆シグナルがそれ以外の部分に含まれるかどうかは、提案手法後のデータに対してスペクトルを算出し、磁気緯度、磁気経度、地磁気状態、季節でビン分けを行い、各ビンの中で電場強度を規格化することで確認できる。実験結果から、DEMETER 衛星データに提案手法を適用するとノイズ分離した電場量らしきものが確認される。雷ノイズに対して、ノイズ除去が可能となる兆候が見られる。しかし、電場量には誤差が生じている。これは、ニューラルネットが周波数特徴を細部まで学習していない、あるいはクラスタリングを行う際に誤った特徴分けを行っているからだと考えられる。ニューラルネット部分の層数変更や半教師ありクラスタリング手法への変更、または学習データの数を増やすことで学習精度の向上が期待できる。

## 5. まとめ

日本は地震大国と呼ばれており、東日本大震災以降、人々の防災意識が高まっている。現在、地震の予測情報として長期予測と短期予測が広く普及しているが、どちらも予測時間が極端である。そこで、電磁気学的な先行現象である夜間 VLF 帯電磁波電場強度減少に工学的な手法を組み合わせることで、地震の短期直前予測が実現すると期待されている。電磁波電場強度を観測する DEMETER 衛星から得られるデータ内には、雷がノイズとして混入しており、このままでは大量にあるデータを活用することができない。そこで、本稿はディープクラスタリングを用いた雷ノイズ

除去手法を提案する。提案手法は、DEMETER 衛星軌道に対して地震前駆シグナルの有無をデータベースで判別した後、観測データから 1.7kHz 帯の DEMETER 衛星データ、電場データ、雷データ、混合データを作成する。学習に用いるデータセットとして電場データ、雷データ、これら 2 データの混合データを使用する。データセットは教師データと検証データに分け、教師データを用いて学習後、検証データによって学習の精度を確認する。テストデータとして、DEMETER 衛星データを使用する。ディープクラスタリングを用いたノイズ除去手法は、作成した 3 種類のデータに短時間フーリエ変換を適用し、各データのパワースペクトルを計算する。次に、2 層の BLSTM とフィードフォワード層で構成されるニューラルネットを用いて、算出した特徴量と正解とするバイナリマスクの差を最小化するように学習する。そして、k-means でクラスタリングし、各周波数ビンがどのデータに帰属するかを識別する。最後に、逆フーリエ変換を用いて分離後の各データを推定する。実験では、提案手法を用いて DEMETER 衛星データを雷ノイズと電場量に分離する。学習が進むにつれて、分離した出力結果の誤差が縮まり、電場量らしきものが確認できる。しかしながら、ノイズの分離精度はまだ不十分である。今後は、ニューラルネットワークの改良やデータ数を増量することで、より精度の高いノイズ除去を目指す。

**謝辞** 本研究は東京大学地震研究所共同利用(2019-Y-地震 3)の援助を受けている。

## 参考文献

- [1] “国土交通省気象庁”  
<http://www.jma.go.jp/jp/quake/>
- [2] 織原義明, 長尾年恭, 地震前兆を科学する, 祥伝社新書(2015)
- [3] Němec, F., O. Santolík, M. Parrot, and J. J. Berthelier, (2008) Spacecraft observations of electromagnetic perturbations connected with seismic activity, *Geophys. Res. Lett.*, 35, L05109.
- [4] Němec, F., O. Santolík, and M. Parrot, (2009) Decrease of intensity of ELF/VLF waves observed in the upper ionosphere close to earthquakes: A statistical study, *J. Geophys. Res.*, 114, A04303.
- [5] Junyuan Xie, Ross Girshick, Ali Farhadi, Unsupervised Deep Embedding for Clustering Analysis, arXiv:1511.06335
- [6] Hershey, J.R.; Chen, Z.; Le Roux, J.; Watanabe, S, Deep clustering: Discriminative embeddings for segmentation and separation, 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)
- [7] “USGS 地震カタログ”  
[http://neic.usgs.gov/neis/epic/epic\\_global.html](http://neic.usgs.gov/neis/epic/epic_global.html)
- [8] DEMETER Microsatellite SCIENCE MISSION CENTER DATA PRODUCT DESCRIPTION
- [9] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [10] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, Vol. 45, No. 11, pp. 2673–2681, 1997.
- [11] TensorFlow  
<https://www.tensorflow.org/>