

波形の類似性を考慮した 大規模月地震データの SOM による可視化システム

後藤 康路^{*1}, 山田 竜平^{*2}, 山本 幸生^{*3}, 横山 昌平^{*1}, 石川 博^{*4}

A system for visualizing large-scale moonquake data considering waveform similarity using SOM

Yasumichi Goto^{*1}, Ryuhei Yamada^{*2}, Yukio Yamamoto^{*3},
Shohei Yokoyama^{*1}, Hiroshi Ishikawa^{*4}

Abstract

Large-scale data of moonquake were obtained from seismometers located on the Moon by the NASA Apollo missions from 1969 to 1977. According to previous analysis of the lunar seismic data, previous studies found that deep moonquakes occur periodically from identical sources at a depth of about 700 to 1200km. The deep moonquakes occurred from identical sources have high similarities among each waveform. This similarity is important to classify the sources and investigate the generation mechanism of moonquakes. From the reason, classification of moonquakes has been processed. We, therefore, develop the web system for visualizing moonquakes considering waveform similarity to progress study of moonquake classification. Our system maps moonquakes data to two-dimensional output space using Self-Organizing Map (SOM). We embed Hadoop in the back-end system to apply SOM to enormous moonquakes data. In this paper, to select a feature for SOM, we evaluate several features based on classified data. Using selected feature, we perform SOM to moonquake data and discuss its result.

Keywords: Moonquake, Visualization, Self-Organizing Map

概要

1969 年から 1977 年の 7 年半の間、NASA のアポロミッションにおいて、月面に設置された地震計を用いた連続観測により、膨大な量の月地震データが取得された。これまでの月地震データの解析より、深発月震は、月深部の約 700-1200km で発生しており、同一の震源から周期的に発生する事が明らかにされている。同一震源から発生する深発月震間において、それらの波形は高い類似性が見られる。この波形の類似性は、震源の分類や、月地震の発生原因の究明において重要な情報となる。そのため、現在に至るまで、人手による月地震の分類が行われている。しかし、膨大な量の月地震データを人手によって分類することは困難である。そこで、本研究では、月地震データの分類研究を促進するため、波形の類似性を考慮して月地震データを可視化するためのシステムを実装する。本システムでは、月地震データを 2 次元空間上へマッピングするため、Self-Organizing Map (SOM) を用いる。膨大な量の月地震データを効率的に処理するため、本システムの処理のバックエンドに Hadoop を用いる。SOM に用いる特徴量を事前に選択するため、月地震データの人手による分類結果に基づいて、Support Vector Machine(SVM) により、月地震を適切に分類可能な特徴量を複数検証する。SVM の分類結果から選択された特徴量を用いて SOM を適用し、その結果について考察する。

1. 背景

NASA のアポロミッションにおいて、Apollo Lunar Surface Experiments Package(ALSEP) と呼ばれる観測装置群が設置された¹⁾。ALSEP を用いた実験のひとつである Passive Seismic Experiment(PSE) は ALSEP に含まれる地震計を用いて月面上で月地震の連続観測を行う実験である。この実験により、約 7 年半に渡る月地震記録が取得された。これらのデータは、長周期地震計の 3 軸成分および短周期地震計の Z 軸成分で構成されており、公開されている²⁾。

*1 静岡大学大学院情報学研究科 (Graduate School of Informatics, Shizuoka University)

*2 国立天文台 RISE 月惑星探査検討室 (National Astronomical Observatory of Japan, RISE project)

*3 宇宙航空研究開発機構宇宙科学研究本部 (The Institute of Space and Astronautical Science)

*4 首都大学東京システムデザイン学部 (Faculty of System Design, Tokyo Metropolitan University)

PSE によって取得された月地震の観測データは、現在も解析が行われており、月地震の発生原因、活動度、および月の内部構造の研究において多くの知見が得られている^{3,4)}。これらの解析の一環として、月地震のクラス分類が行われており、波形から予測された発生深度や発生要因をもとに、各月地震に深発月震、浅発月震、隕石衝突などのラベルが付与されている。特に、深発月震は同一震源から周期的に発生することが分かっているため^{5,6)}、月-地球間に働く潮汐力と関連して、その発生原因について研究が進められている^{7,8)}。これらの研究を通して、深発月震の震源にはラベルが付与されている。これらの付与は、人手によって行われ、月地震イベントカタログとして公開されている。この月地震イベントカタログは、現在でも多数の研究に基づいて更新が行われている⁹⁾。しかし、PSE によって取得された月地震のデータは、膨大な量である。そのため、未だにラベルが付与されていない月地震や人手ではラベルの付与が困難な月地震が存在する。また、人手によるラベルの付与の正確性の検証が困難であるという課題がある。そこで、本研究では、コンピュータを用いた月地震解析のファーストステップとして、月地震データの解析・可視化システムを構築する。本システムは、月地震の波形の類似関係を可視化し、人手により付与されたラベルと比較を行うことで、ラベルが付与されていない月地震の解析の補助、付与されたラベルの検証、および未知クラスの発見補助を目的とする。

本研究では、月地震解析基盤と、解析結果の可視化インターフェイスを提案する。月地震解析基盤では、Self-Organizing Map(SOM)¹⁰⁾を用い、高次元データである月地震波形を、波形の類似度に基づいて低次元空間へのマッピングを行う。また、SOM を適用するため、SOM に用いる特徴量の選択を事前に行う。そのため、教師あり機械学習手法のひとつである Support Vector Machine(SVM)を用いて、特徴量ごとに分類結果の評価を行い、最も人手によるラベルの付与に近いと評価された特徴量を用いる。可視化インターフェイスでは、月地震解析基盤で得られた SOM の結果と、人手による分類結果と合わせて提示する。解析結果と人手による分類結果を比較することにより、付与されたラベルの正確性の検証や、人手による分類結果に反するような性質を示す波形の発見を促進する。

2. 提案システム

本研究で提案するシステムの構成図を図 1 に示す。Web インターフェイスは Web サーバ、月地震解析基盤はクラスタによって構築されている。月地震データセットは、約 1TB の大規模な時系列データであり、13,058 件の月地震を含んでいる。そのため、最終的に全データの解析を目的とする本研究では、SOM や SVM の計算量が膨大になると考えられる。このような課題に対して、分散処理の導入が考えられる。しかし、分散処理プログラムの開発コストは非常に高価である。そこで本研究では、分散処理の導入を容易にするため、大規模データ処理基盤である Hadoop¹¹⁾と、Hive¹²⁾を用いる。ユーザの要求を受け、Hadoop クラスタ内の多数のコンピュータを用い、月地震データの管理、SOM、および SVM 処理を行う。ユーザへの可視化結果の提示には、HTTPD を介した Web インターフェイスを用いる。可視化インターフェイスでは、ユーザの要求に応じて、SOM の結果に基づいた Web ページを生成する必要があるため、リアルタイムなデータ応答が必要になる。しかし、Hadoop はリアルタイムなデータ応答に適していない。そのため、予め Hadoop によってデータ処理を行い、その結果を RDBMS に蓄積するという形式をとる。Hadoop と RDBMS 間のデータ転送には Sqoop¹³⁾を用いる。

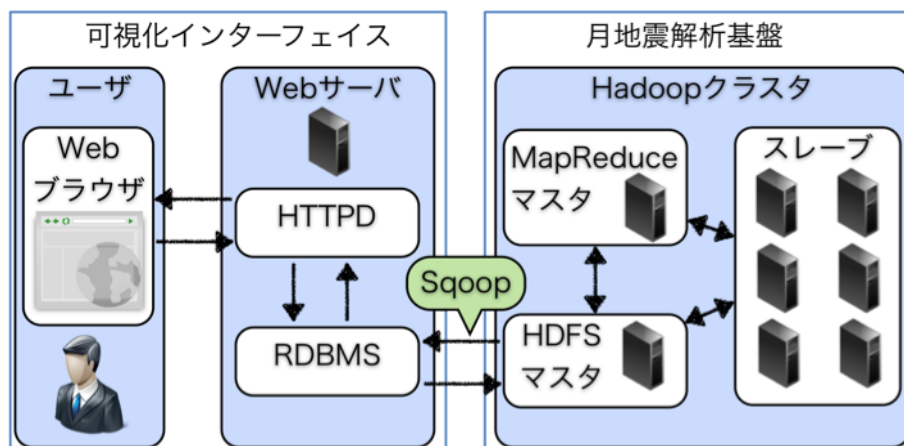


図 1 提案システムの構成図

2.1. データ管理

月地震解析基盤で用いる Hadoop は、分散ファイルシステム HDFS と、分散処理フレームワーク MapReduce により構成されている。大規模月地震データを HDFS で管理し、MapReduce によってデータ処理を行うことによって、分散並列処理によるデータ解析の効率化を図る。また、Hive を用いることにより、これらのファイルを SQL ライクなドメイン固有言語を介して MapReduce 操作を行うことができる。Hive はユーザ定義関数の定義ができるため、MapReduce を用いたプログラムを直接作成する場合と比較して、高度なデータ処理を容易に記述および実行できるようになる。また、HDFS に格納された圧縮ファイルに対するデータ操作もサポートしており、MapReduce 処理におけるコンピュータ間の通信のオーバーヘッド削減や、専有するストレージ容量の削減の効果がある。

月地震データの管理では、RCFile¹⁴⁾形式を用いる。RCFile は、Hadoop 上でのデータ格納形式のひとつである。列指向データファイル形式という特徴を持つため、列単位でのデータの管理およびデータ操作を効率的に行うことができる。月地震データは、時系列データとして保存されるため、列指向データファイル形式が適していると考えられる。また、RCFile と合わせ、高速データ圧縮手法である Snappy 圧縮¹⁵⁾を用いる。これらの手法により、TSV 形式で約 1TB のデータセットに対して、HDFS 上において約 1/5 のサイズに圧縮し、データ管理を行った。

2.2. SOM

教師なし機械学習手法のひとつである SOM は、入力データの類似性に基づいて低次元の出力空間へ入力データをマッピングする手法である。各データは、出力空間上に配置されたセルにマッピングされる。各セルは、入力データと同じ次元の重みベクトルを持つ。SOM は、入力データの傾向に基づいてセルの重みベクトルに学習し、入力データを最も類似したセルに配置することで、低次元空間へのマッピングを行う。入力データ x_n に最も類似したセルを返す関数 $c(x_n)$ を式 (1) に示す。

$$c(x_n) = \arg \min_p d_{pn} \quad (1)$$

ここで、 d_{pn} は、 x_n とセル p の距離を表す。

学習時には、一度出力空間上に入力データをマッピングし、セルの周囲に配置された入力データを用いて、重みベクトルの更新を行う。学習は、学習係数 σ を 1 から 0 になるまで単純減少させながら行う。セル r_i , r_j 間の影響度を表す近傍関数 $h(r_i, r_j)$ を式 (2)、学習の正則化項 γ_p を式 (3)、重みベクトルの更新式を式 (4) に示す。

$$h(r_i, r_j) = \exp\left(-\frac{\|r_i - r_j\|^2}{2\sigma^2}\right) \quad (2)$$

$$\gamma_p = \frac{1}{\sum_i^N h(p, x_i)} \quad (3)$$

$$m_p(t+1) = \gamma_p \sum_i^N h(p, c(x_i)) x_i \quad (4)$$

$m_p(t+1)$ は、セル p の更新後の値、 N は入力データ数を表している。これらの処理により、類似したデータを同じセルに、セル同士の距離に近いほど、セル間の類似度が高くなるようにデータを配置する。高次元特徴量を持つデータを各データ間の類似性を保ちながら低次元空間にマッピングすることで、ユーザはそれらのデータを容易に閲覧可能になる。

本システムでは、月地震データが線形分離不可能である場合を考慮し、データが線形分離不可能な場合にも適用可能なカーネル SOM¹⁶⁾を用いる。一般的な SOM は、距離関数を用いてデータ間の類似度を計算するのに対して、カーネル SOM は、カーネル法を用いてデータ間の類似度を計算する。カーネル法は、非線形データに対するデータ解析で広く用いられている手法であり、データを高次元へ非線形写像することで、解析手法を非線形データに対して拡張する。実際は、高次元での解析は膨大な計算量を伴うため、写像後の空間における内積を定義するカーネル関数を用いて類似度の計算を行う。カーネル関数を用いる場合、セルの重みベクトルを求めることができないため、直接セルと各データの距離を求める。式 (5) にセル p とデータ x_i の距離の式を示す。ここで、 $K(x_i, x_j)$ はカーネル関数を表す。

$$\begin{aligned}
d_{pn} = & K(\mathbf{x}_n, \mathbf{x}_n) \\
& - 2\gamma_p \sum_i^N h(\mathbf{p}, c(\mathbf{x}_i)) K(\mathbf{x}_n, \mathbf{x}_i) \\
& + \gamma_p^2 \sum_i^N \sum_j^N h(\mathbf{p}, c(\mathbf{x}_i)) h(\mathbf{p}, c(\mathbf{x}_j)) K(\mathbf{x}_i, \mathbf{x}_j)
\end{aligned} \tag{5}$$

3. SVM を用いた特徴量の評価と選択

月地震データセットを、カーネル SOM を用いて可視化する際に、各データ間の類似度を計算するために用いる特徴量を選択する。本研究では、月地震データの人手による分類結果に基づいた特徴量の選択を行うために教師あり機械学習手法のひとつである SVM を用いる。SVM は、教師あり機械学習手法のひとつであり、クラス分類によって予測ラベルが計算される。そのため、SVM によるクラス分類の結果と人手による月地震分類の結果を比較し、定量的に月地震の特徴量を評価することができる。本実験では、人手による月地震の分類結果を SVM のトレーニングデータとし、複数の特徴量について SVM を適用し比較・考察を行う。

今回の実験では、システムの適用例として、月地震データセットから波形を抽出し、実験用のデータセットとする。用いる波形は、人手により深発月震と分類された月地震のうち、震源が A1, A6, A7, A8, A9, および A10 と分類されている月地震波形 551 件である。これらの波形を選択した理由として、震源位置が良く決められており波形と震源との関連が明瞭である点と、各震源に対して地震波形のサンプル数が充分であるため、特徴量の評価が容易であるという点が挙げられる。特に、実験で使用した 551 件は S/N 比が高く、比較的 S 波が明瞭に識別できるものを選択している。各震源に含まれるデータ数の内訳を、以下の表 1 に示す。各波形データは、月地震の S 波を中心に、前後約 10 分程度になるように調整し、トレンド引きやノイズ除去といった前処理を適用したデータを利用する。

表 1 実験に用いる月地震データ数の内訳

震源クラス	A1	A6	A7	A8	A9	A10	合計
要素数	204	54	53	55	106	79	551

本実験において用いる特徴量算出手法を以下に示す。

- Power Spectral Density (PSD)
- フレーム毎に区切った波形の PSD
 - (フレーム長: 波形の 1/16, シフト長: 波形の 1/32, 窓関数: ハニング窓)
- PSD のエンベロップ
- フレーム毎に区切った波形の PSD エンベロップ

PSD は、周波数成分毎に振幅の強度を計算したものであり、時系列データの時間相関分析などに用いられる。PSD から特徴量を算出する手法のうち、広く用いられる手法として、フレーム分割やローパスフィルタによるエンベロップを用いる手法が挙げられる。本実験では、これらの特徴量算出手法を用いて特徴量を複数の波形形式から抽出し、評価を行い、適切な特徴量を選択する。波形形式は、直接波形データの値を用いる場合と、波形データのエンベロップを用いる場合の 2 種類である。本実験では、4 種類の特徴量算出手法と 2 種類の波形形式の組み合わせである 8 特徴量について評価を行う。図 2 に、8 特徴量の導出手順を示す。本実験において、エンベロップは移動平均に基づいて算出する。

今回はカーネル関数として線形カーネルを用いる。評価指標としては、クラス毎に適合率、再現率、および F 値を用い、10 分割交差検定によって検証を行う。適合率は、予測結果に対して、正しく予測できた割合を示す値である。再現率は、正解に対して、予測結果が網羅できた割合を示す値である。F 値は、適合率と再現率の調和平均である。

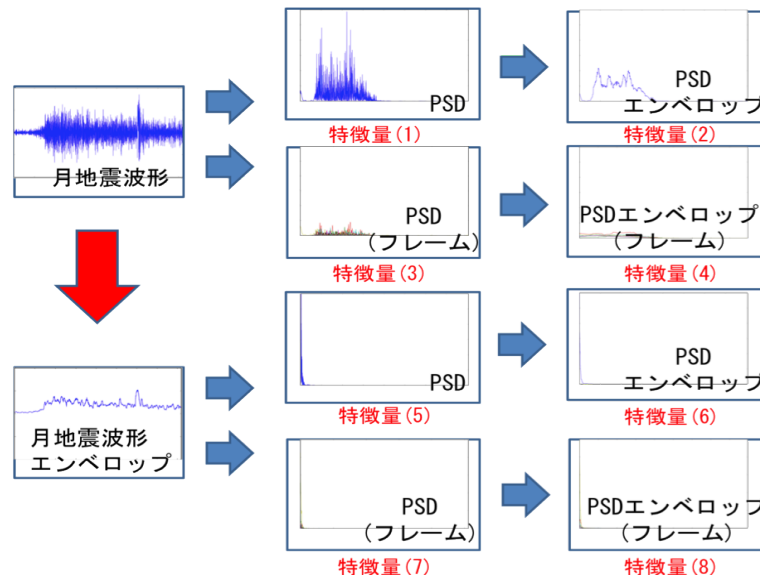


図2 特徴量抽出手順

図3、図4および図5に月地震波形を用いた場合の実験結果、図6、図7および図8に月地震波形のエンベロップを用いた場合の実験結果を示す。実験結果より、月地震波形を用いた場合と月地震波形のエンベロップを用いた場合の両方において、PSDを特徴量に用いた場合のF値の平均値が最も高く評価されたため、PSDがより適切な特徴量であることが確認された。また、図5の月地震波形を用いた場合と図8の月地震波形のエンベロップを用いた場合の実験結果の比較より、最もF値の平均が高い特徴量は、月地震波形を用いた場合のPSDである。そのため、次章では、月地震波形を用いた場合のPSDに基づいてカーネルSOMを適用する。

ここで、最もF値の平均値が高い、月地震波形を用いた場合のPSDの結果に着目する。各震源の分類精度を比較すると、A1のクラス分類精度は、他の震源のデータと比較して高い。この結果について、A1を震源とする波形は、他の震源とは異なる性質を含んでいるためF値が高い値になったと考えられる。また、A8について、他の震源の適合率と比較して、A8の適合率は低い。このことから、A8を震源とする月地震データが含む性質を、A8以外の震源の月地震データの一部も含んでいる可能性があると考えられる。しかし、このようなクラス分類の結果のみに基づいて、月地震に対する検証や考察を行うのは困難である。そこで、次章では、カーネルSOMを用いた可視化手法を適用することにより、月地震データの検証を行う。

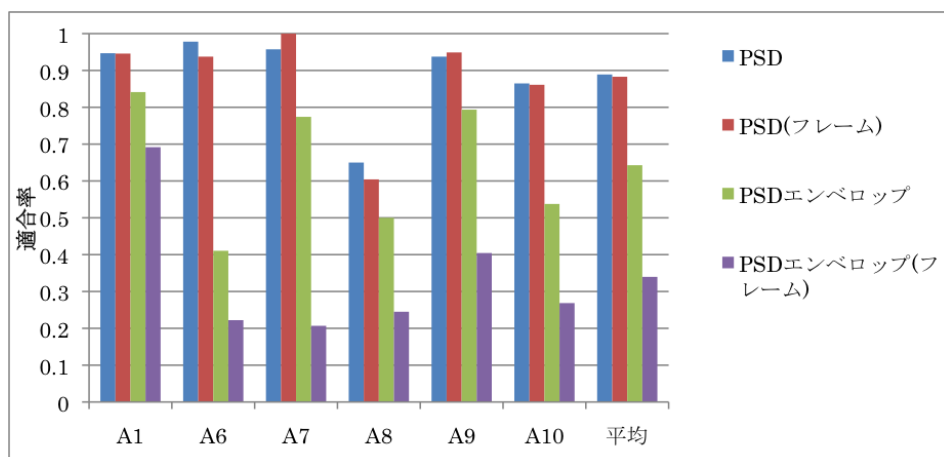


図3 特徴量選択のための評価実験結果（月地震波形を用いた場合）：適合率

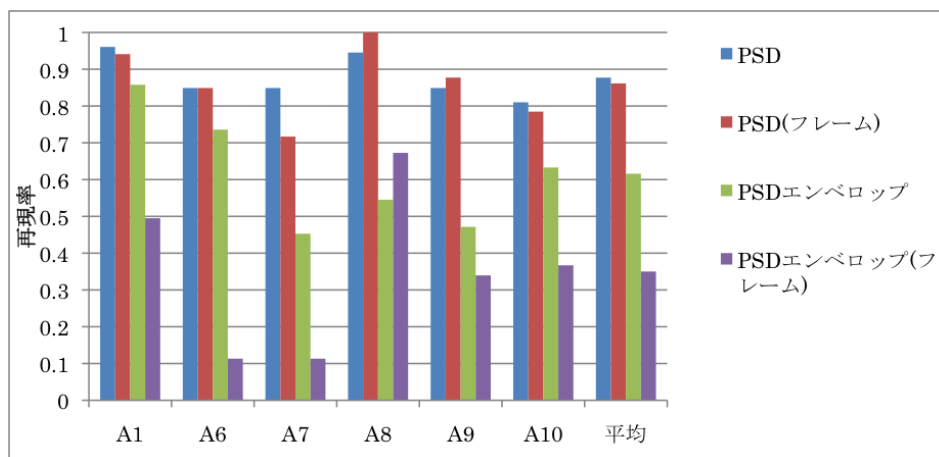


図4 特徴量選択のための評価実験結果(月地震波形を用いた場合): 再現率

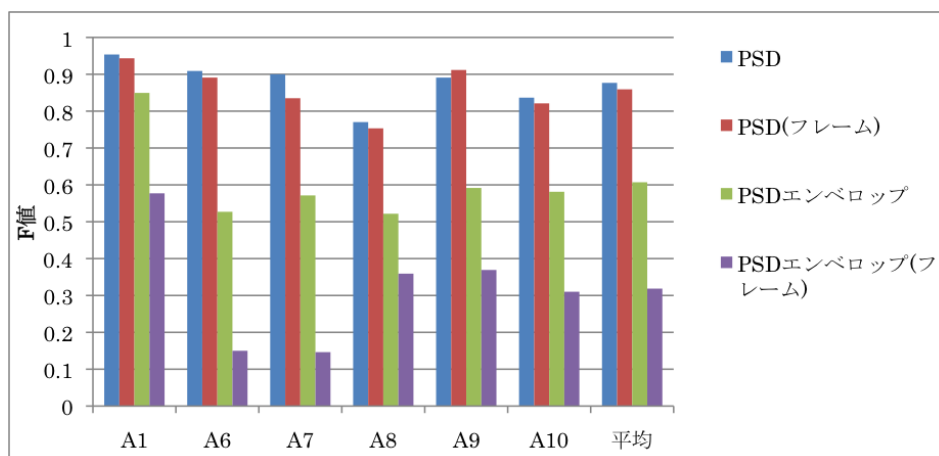


図5 特徴量選択のための評価実験結果(月地震波形を用いた場合): F 値

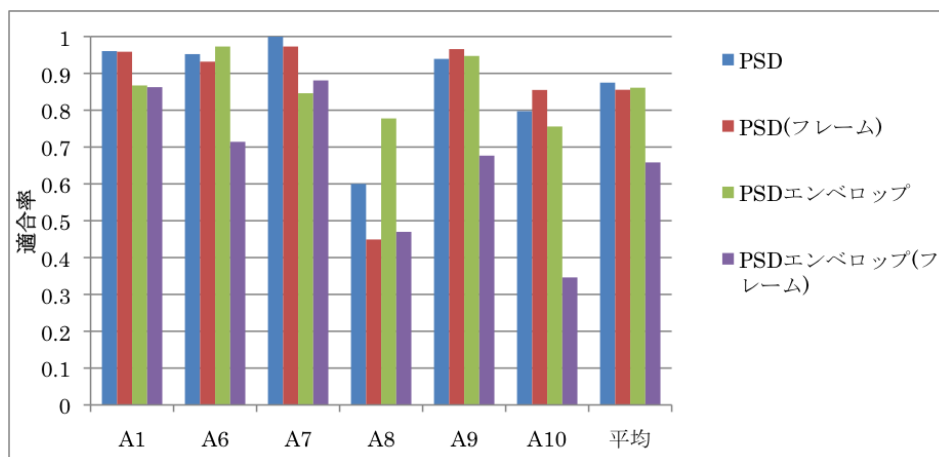


図6 特徴量選択のための評価実験結果(月地震波形のエンベロップを利用した場合): 適合率

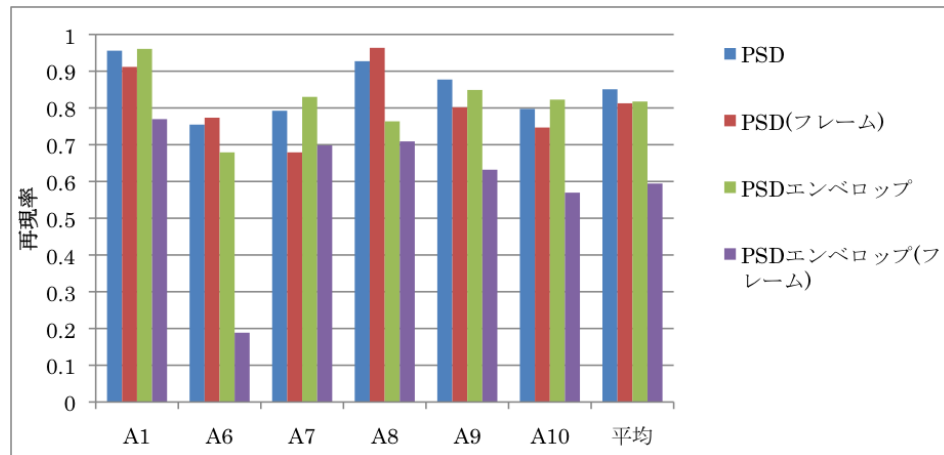


図 7 特徴量選択のための評価実験結果（月地震波形のエンベロップを利用した場合）：再現率

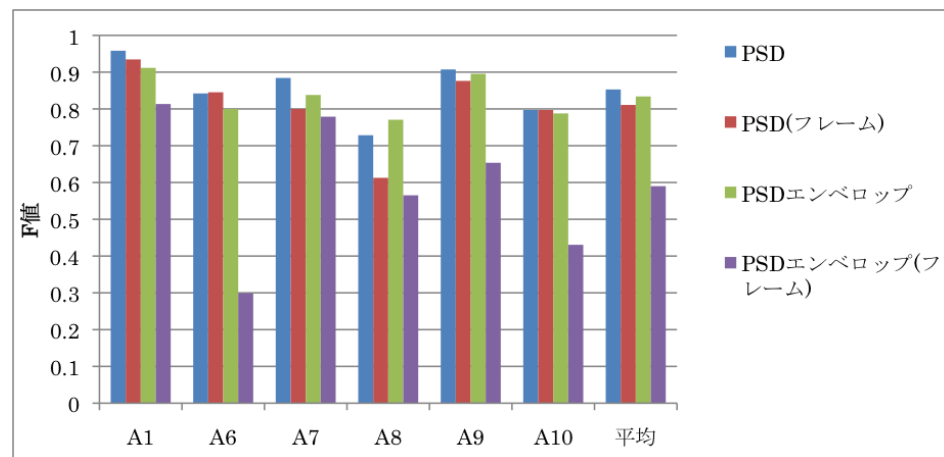


図 8 特徴量選択のための評価実験結果（月地震波形のエンベロップを利用した場合）：F 値

4. カーネル SOM による月地震の可視化例

月地震解析基盤において、月地震データに対してカーネル SOM を適用する。また、SOM の結果に基づいて、可視化インターフェイスを用い、人手によって付与されたラベルが他のデータとどのように類似しているかを可視化する。カーネル SOM に用いる特徴量は、前章の実験において最も F 値の平均値が高い、月地震波形を用いた場合の PSD を用いる。カーネル SOM で用いるカーネル関数は、線形カーネルを用いる。カーネル SOM を適用するデータは、前章の実験に用いた月地震データと同様である。本章では、カーネル SOM のパラメータについて、複数のパラメータで実験を行い、適切と考えられる値を設定した。具体的なカーネル SOM のパラメータは以下の通りである。

- 出力空間：トラス空間
- セルの数：25(5 × 5)
- 学習のイテレーション回数：10 回

図 9 は、カーネル SOM の適用例である。各セルの円グラフは、セルに配置された月地震について、人手によって付与されたラベル毎に色分けし、内訳を表示している。

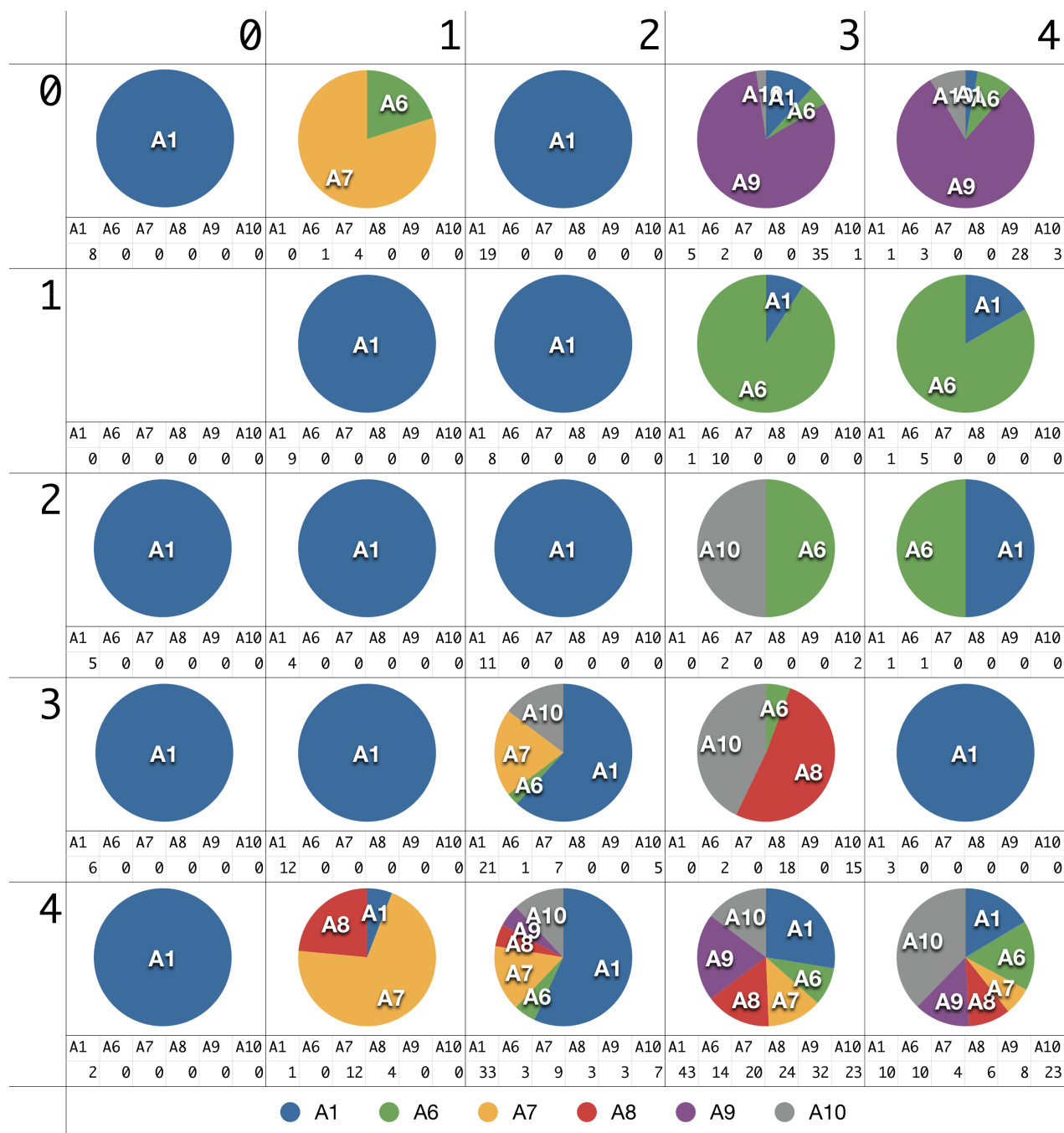


図9 深発月震に対するカーネル SOM 適用例

前章の結果において、F 値が高かった震源 A1 のデータ (図9の円グラフにおいて青色) に着目すると、可視化結果では、全 25 個のセルのうち A1 が震源である月地震データのみを含むセルが 11 個存在し、非常に広範囲のセルを A1 のみで専有している。この結果より、震源 A1 から発生したとされる月地震データは、他の震源との類似度が低く特徴的な性質の波形を含むことが分かる。震源 A6 のデータ (図9の円グラフにおいて緑色) に着目すると、セル 1 行 3 列目周辺に全体の約半数のデータが集まり、ほぼセル専有している。震源 A7 のデータ (図9の円グラフにおいて黄色) は、セル 4 行 1 列目周辺に集まり、セル 4 行 1 列目を、ほぼ専有している。震源 A8 のデータ (図9の円グラフにおいて赤色) は、セルを専有してはいないものの、セル 3 行 3 列目およびセル 4 行 3 列目に集中している。震源 A9 のデータ (図9の円グラフにおいて紫色) は、セル 0 行 3 列目周辺に集まっており、特にセル 0 行 3 列目と 0 行 4 列目を、ほぼ専有している。震源 A10 のデータ (図9の円グラフにおいて灰色) も、A8 と同様にセルを専有してはいないものの、セル 3 行

目 3 列目およびセル 4 行目 4 列目に集中している。セルの専有の有無は、3 章の実験において、A8 と A10 の F 値が比較的良かったことと対応していると考えられる。

また、右下のセル 4 行 3 列目とセル 4 行 4 列目において、多数の震源のデータが集まっていることが分かる。前章の月地震波形を用いた場合の PSD でのクラス分類結果において、A8 と誤分類された波形は全 28 件であり、その内訳は正解震源 A1 が 4 件、A6 が 3 件、A7 が 4 件、A9 が 7 件、A10 が 10 件である。これらの月地震に着目すると、セル 4 行 3 列目に 20 件、セル 4 行 4 列目に 8 件の月地震が集まっていることが分かった。これらの月地震の波形を図 10、図 11 に示す。また、図 12 にセル 4 行 1 列目に配置された A8 を震源とする波形を示す。分類結果の正誤に関わらず、セル 4 行 3 列目に配置された波形は、波形の振幅が小さいという共通点がある。そのため、波形の性質がノイズに埋もれ、分類に失敗した結果、A8 と誤分類されたものと考えられる。A8 の適合率は、A8 の月地震データが正しく分類された数 TP、A8 以外の月地震データが A8 と誤分類された数 FP より、 $TP/(TP+FP)$ として求められるため、誤分類の増加によって A8 の適合率が低下したと考えられる。

同様に、A10 と分類された月地震波形に着目する。図 13 と図 14 は、セル 4 行 3 列目に配置された波形である。また、図 15 に、セル 2 行 3 列目に配置された A8 を震源とする波形を示す。これらの波形についても、分類結果の正誤に関わらず、セル 4 行 3 列目に配置された波形は、波形の振幅が小さいという共通点がある。そのため、セル 4 行 3 列目に配置されたデータには、波形の性質がノイズに埋もれ、分類が困難である波形が集まったと考えられる。

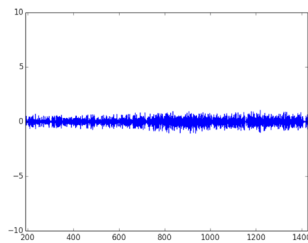


図 10 分類結果 A8
正解震源 A6 の波形
(セル 4 行 3 列目に配置)

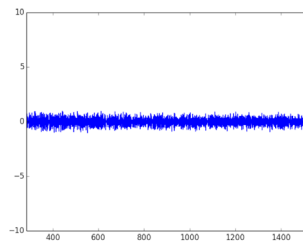


図 11 分類結果 A8
正解震源 A8 の波形
(セル 4 行 3 列目に配置)

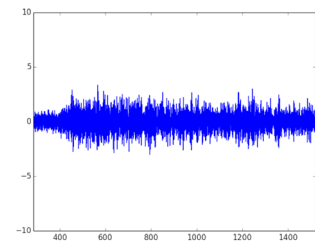


図 12 分類結果 A8
正解震源 A8 の波形
(セル 4 行 1 列目に配置)

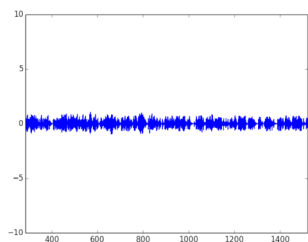


図 13 分類結果 A10
正解震源 A8 の波形
(セル 4 行 3 列目に配置)

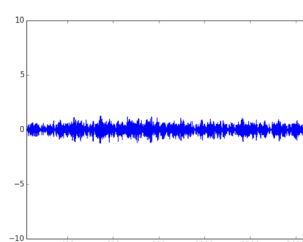


図 14 分類結果 A10
正解震源 A10 の波形
(セル 4 行 3 列目に配置)

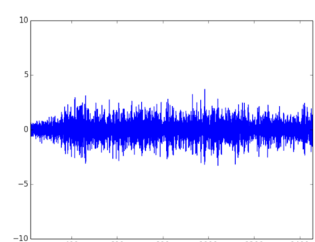


図 15 分類結果 A10
正解震源 A10 の波形
(セル 2 行 3 列目に配置)

5. まとめ

本研究では、波形の類似性を考慮した月地震データの解析・可視化システムの実装を行った。大規模データに適用するために、月地震解析基盤に Hadoop を用いた。可視化手法としてカーネル SOM を用い、その結果を提示する可視化インターフェイスを実装した。また、システムの適用例として、一部の深発月震の震源分類について、教師あり機械学習手法である SVM の結果と比較し、考察を行った。SOM に基づいた可視化結果より、震源 A1 について、他の震源と異なる性質を持つ波形が多いことや、震源 A8 について、ノイズにより波形の持つ性質が埋没し、そのままでは他の震源との分類が困難であることが分かった。しかし、このように人手による分類結果と機械的な分類結果との比較を円滑

に行うことができるという点において、本システムは、今後、人手による月地震分類結果の検証や、未分類とされている月地震の分類を行っていく上で、有効であると考えられる。

今後は、他のカーネル関数を用いた精度向上や、全月地震データセットについて、他の観測点で得られた月地震データ、震源位置および未知クラスの月地震波形を考慮した可視化を進めていく予定である。

参考文献

- 1) N. Toksöz, A. Dainty, C. Solomon and K. Anderson, “Structure of the Moon,” *Reviews of Geophysics and Space Physics*, 12, 1974, pp. 539-565.
- 2) 山田竜平, 山本幸生, 桑村潤, 中村吉雄, “アポロ月地震データ公開システムの開発”, 宇宙科学情報解析論文誌 宇宙航空研究開発機構研究開発報告, 第一号, 2012, pp. 121-131.
- 3) Y. Nakamura, G. Latham and J. Dorman, “Apollo lunar seismic experiment –final summary” , *Journal of Geophysical Research*, 87, 1982, pp. A117-A123.
- 4) P. Lognonné, J. Gagnepain-Beyneix and H. Chenet, “A new seismic model of the moon: Implication for structure, thermal evolution and formation of the moon” , *Earth and Planetary Science Letters*, 211, 2003, pp. 27-44.
- 5) D. R. Lammlein, “Lunar seismicity and tectonics” , *Phys. Earth Planet. Inter.*, 14, 1977, pp. 224-273.
- 6) R. C. Bulow, C. L. Johnson, B. G. Bills and P. M. Shearer, “Temporal and spatial properties of some deep moonquake clusters” , *Journal of Geophysical Research: Planets*, 112.E9, 2007.
- 7) H. Araki, “Focal processes of deep moonquakes” , *J. Geod. Soc Japan*, 47, 2001, pp. 508-513.
- 8) C. C. Hon and N. Toksöz, “Tidal stresses in the Moon” , *Journal of Geophysical Research*, 83.B2, 1978, pp. 845-853.
- 9) Y. Nakamura, G. Latham, J. Dorman and J. Harris, “Passive seismic experiment long-period event catalog” , *Galveston Geophysics Laboratory Contribution*, 491, 2008.
- 10) T. Kohonen, “The self-organizing map” , *Proceedings of the IEEE*, 78.9, 1990, pp. 1464-1480.
- 11) “<http://hadoop.apache.org/>” .
- 12) “<http://hive.apache.org/>” .
- 13) “<http://sqoop.apache.org/>” .
- 14) Y. He, R. Lee, Y. Huai, Z. Shao, N. Jain, X. Zhang and Z. Xu, “RCfile: A fast and space-efficient data placement structure in mapreduce-based warehouse systems” , *The IEEE International Conference on Data Engineering (ICDE)*, 4, 2011, pp. 1199-1208.
- 15) “<http://code.google.com/p/snappy/>” .
- 16) P. Andras, “Kernel-kohonen networks” , *International Journal of Neural Systems*, 12, 2002, pp. 117-135.