

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 19 日現在

機関番号：13801

研究種目：挑戦的萌芽研究

研究期間：2013～2014

課題番号：25540038

研究課題名(和文) 考古学的アプローチによる埋没ビッグデータの再センシング基盤技術

研究課題名(英文) Re-sensing Infrastructure of Hidden Big-data Based on Archaeology Approach

研究代表者

横山 昌平 (Yokoyama, Shohei)

静岡大学・情報学研究科・講師

研究者番号：20443236

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：1969年から77にかけて月の大規模な地震観測がNASAのアポロ計画の一部として行われた。本研究ではそのデータを要因・震源別で自動分類する手法を実現し、それを人手による分類と対比できる形で可視化する事を目指した。我々のシステムは自己組織化マップ(SOM)に基づいており、高速な処理を可能とするためにHadoop、Spark上で並列実行する基盤技術を確立した。また波形特徴量に関して、月の地震波データに対して最適な特徴量抽出手法を検討し、それに基づきSOMの出力である二次元グリッドスペースに可視化するシステムを実現した。本成果は月地震に関する研究者が新たな知見を獲得するための重要なツールとなる。

研究成果の概要(英文)：Large-scale seismic data were obtained from seismometers located on the Moon by the NASA Apollo missions from 1969 to 1977. We developed a web system for visualizing moonquakes considering the waveform similarity to progress study of moonquake classification. Our system maps moonquakes data to two dimensional output space using Self-Organizing Map (SOM). We implemented parallel-SOM based on Hadoop and Spark as back-end of moonquake analysis. We also evaluated several features based on classified data and discovered effective features from seismic data stream. Our visualization system can help researchers of moon quake to discover new knowledge of moon.

研究分野：データ工学

キーワード：宇宙科学データ ソーシャルデータ

1. 研究開始当初の背景

本研究計画は、時代に埋もれたビッグデータを最新のデータマイニング技術を使って再処理するための基盤技術を研究開発する。具体的には我々は国立天文台およびJAXAの研究者と連携し、米国アポロ計画で月に設置された複数の地震計のデータを用いる計画である。通常のセンサストリームからのデータマイニングと決定的に異なるのは、「正解」が不明であるという点である。アポロの地震計はすでに運用を終えており、月の地震(月震)に関する新たなデータを取得するならば、最低でも3台の人工衛星を月に送り込まなければならず現実的ではない。すなわち、我々は40年前に観測され磁気テープに保存されていたビッグデータを発掘し、センサ精度・通信・データ保存で生じたノイズを修復したデータでしか、月の状況を知る事ができない。これを我々は考古学的アプローチと名付け、再センシングの為の新たな技術基盤の創造をめざす。

2. 研究の目的

本研究計画はビッグデータの処理という新しいトピックに対して、50年前の古いデータの適用を検討する点で特徴的である。また実際に月の地震を対象とする固体地球科学分野の研究者を連携研究者に迎え共同研究スキームを構築する事で、実際の観測データを使った実用的な手法の実現を目指す点で独創的である。月震データはこれまでアポロ計画によってのみ観測されている。この観測からすでに50年経っているが、このビッグデータに対して固体地球科学と情報科学の共同研究はこれまでに例が無く、本研究計画は萌芽的な課題と言える。

現在の技術から考えると、今観測されたデータは今後50年経とうが劣化はしないと考えられるが、今から50年前は、現在とは情報技術が全く異なっている。その為、その当時のデータに、現在のデータマイニング技術を、そのまま適用する事は困難であり、我々はそこに挑戦的な研究課題があると考えている。

例えば、データは磁気テープに保存されていたため、テープの切り替え場所のノイズを考えなければならない。他にも、月からの通信技術が未熟であったため、通信の為のヘッダ情報が、センシングデータの一部を上書きする形で付与されており、その点の補完等も考えなければならない。また、提案手法では、その当時のビッグデータとして代表的な宇宙科学分野のセンシングデータを主なターゲットとしているが、そのデータ特有の問題もある。例えば地球における地震の観測であるならば、地震計の精密なキャリブレーション

ンが行えるが、月ではそれは叶わない。また地震計もこれまでに6台しか設置されておらず、すでに運用を停止しているため、我々はすでに得られているデータのみから新たな知見を発見する事を考えなければならない。月震波形の発見と分類(隕石の衝突が原因なのか、あるいは月の内部での変動が原因なのか等)はテキサス大学のNakamuraによってなされているが、ヒューリスティクスに基づいており、また分類不明な月震が半数に及ぶ。そこで機械学習などのデータマイニング手法を使い、再分類と新しい月震の発見を目指す。

一般的にセンシングデータに対し、そのような処理を行う場合、すでに分類されている項目を正解データとみなした教師あり学習を行う事が考えられる。しかし月震の場合、月の内部構造が解明されている訳ではないため、これまでの分類を正解としてしまうと、新たな知見の発見を逃す可能性が高い。そこで提案手法では、まず教師なし学習とクラスタリングにより、過去の月震分類から有意差を見つけるところ端緒に研究を行う。

この様に、過去にセンシングされ埋没していたデータならではの課題が数多くあり、その解決を目指す事が本研究のチャレンジである。

3. 研究の方法

本研究は主対象のデータを月の地震波としているが、一般化するとリアルタイムの観測では無い時空間データストリームに対する、データマイニング手法の確立である。ただし本研究は「正解データ」を得る事が出来ないという事が大前提となる点で特徴的である。最初に機械学習により月地震を観測した波形データの分類を試みた。

月地震の原因は大きくは浅発地震、深発地震、隕石衝突等に分類され、さらには震源域別に分類される。これは従来人手によって行われている。月の内部構造は地球程解明が進んでおらず、地球のようにピンポイントに震源を判別できる訳ではなく、従来は科学者の知識の蓄積により、分類されてきた。今回の萌芽的アプローチはそれを正解として学習し分類器を作るのではなく、正解無しに分類を行い、人手による分類をどれくらい再現できるか、また再現できなかった部分に、科学的な新たな知見があるのかを可視化により明らかにするというアプローチを採った。

そこで教師なし学習手法でかつ可視化との親和性が高い Self-Organized Map (SOM) を基礎として手法を構築した。SOMは高次元データを低次元空間へ写像する手法であり、高次元の波形特徴量を差異に基づいて効果的に二次元空間で可視化する事ができる。

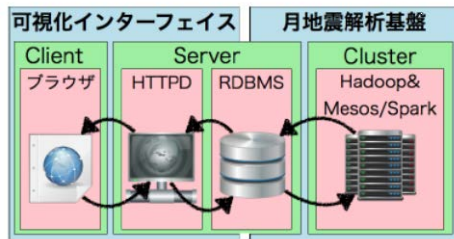


図 1 システム全体像

しかしながら、SOM は出力マップのサイズを入力データに合わせて適切に設定する必要がある。そこで本研究では Growing Hierarchical SOM という手法を改良して適用した。

また、月地震データで得られた、「正解を知らぬデータ」に対する解析技術・ノウハウを、同様に正解を得るのが難しい、ソーシャルビッグデータへの適用も行った。

4. 研究成果

我々の対象とする月地震データセットは、約 1 TB の時系列データであり、これまで人手により 13,058 件の月地震が観測されている。宇宙科学分野における目標は、この地震波のデータを解析する事により、月の組成や成り立ちを解明する事にあるが、我々計算機科学としての目的は、それに加えこのビッグデータに対してどのような効率的処理を実装するかという点にある。

GHSOM の計算量は膨大であり、分散処理をさせる事によりスケーラビリティの確保を目指した。分散処理基盤として代表的な手法には MapReduce があるが、対象データの処理は基本的にはイテレーションであり、ジョブ毎にオーバーヘッドが発生する MapReduce は適さない。そこで Spark を用いた。Spark は、ユーザからの要求に従い、PC クラスタ上の Mesos を用いて GHSOM の処理を行う。Mesos は、効率的なリソース管理を行うためのリソースマネージャであり、Spark をはじめとする分散処理システムにおいて広く用いられている。ユーザへの可視化結果の提示には、Web サーバ を介した Web インタフェースを用いる。可視化インタフェースではユーザとのインタラクションを重視し、レイテンシの削減の為、処理結果のキャッシュとして RDBMS を補助的に用いた。図 1 にシステムの全体像を描く。

図 2 は我々の実現したシステムにおいて、月地震の波形を分類し可視化した一例である。SOM によって地震波の多次元データを二次元のグリッドにマップした。そして各グリッドにおいて、Nakamura によってラベル付けされた、震源域の割合を円グラフで表示している。すなわち、もし 1 つの



図 2 : 可視化インタフェース

セルに、同じ震源とラベル付けされている波形データが集まった場合は、人手による分類が、波形特徴量により把握できるという証となり、そこに少数の別の震源で起こった月地震の波形が含まれていた場合、その波形は、人手による分類に関わらず、波形の特徴からは、そのセルにおいてドメスティックな震源の地震はと類似している事を示している。

しかしながら、前述したように、正解を得る事は出来ないデータのため、これを可視化する事で、月地震が専門の科学者が、あらたな知見を発見するためのツールとして活用を期待している。

このように、正解のないデータ、データ取得時の状況を把握できないデータを可視化する事により、分析に役立てる方法論を確立し、そのためのツールを実装した事が、本研究の成果である。月地震に関しては、研究機関の 2 年間に於いて、学術論文誌 1 報、国際会議 1 報、国内研究会 2 報の発表を行った。加えて、本研究の成果を活用してソーシャルデータに適用した研究での成果も多く得られた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 5 件)

- ① 大森 雅己, 廣田 雅春, 石川 博, 横山 昌平, "ソーシャルメディア上から収集したジオタグに基づく地理的特徴の抽出と評価", 情報処理学会論文誌データベース (TOD), Vol. 8, No. 1, pp. 1-16, 2015. 03. 30, 査読有
- ② 大羽 洋隆, 廣田 雅春, 石川 博, 横山 昌平, "ジオタグ付き写真を用いた土地被覆の分析と可視化", 日本データベース学会和文論文誌, Vol. 13-J, No. 1, pp. 52-57, 2014. 10, 査読有
- ③ 白井 元浩, 廣田 雅春, 石川 博, 横山 昌平, "ジオタグ付き写真を用いた関心領域

と撮影スポットの発見”, 電子情報通信学会論文誌 D, Vol. J97-D, No. 4, pp. 835-844, 2014. 04. 01, 査読有

- ④ 後藤 康路, 山田 竜平, 山本 幸生, 横山 昌平, 石川 博, “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”, 宇宙航空研究開発機構研究開発報告 宇宙科学情報解析論文誌, Vol. 3, pp. 137-146, 2014. 03. 31, 査読有

- ⑤ Ágnes Bogárdi-Mészöly, András Rövid, Hiroshi Ishikawa, Shohei Yokoyama, Zoltán Vámosy, “Tag and Topic Recommendation Systems”, Acta Polytechnica Hungarica, Journal of Applied Sciences, Vol. 10, No. 6, pp. 171-191, 2013, 査読有

[学会発表] (計 8 件)

- ① Masaharu Hirota, Masaki Endo, Shohei Yokoyama, Hiroshi Ishikawa, “Visualizing Shooting Spots using Geo-tagged Photographs from Social Media Sites”, 15th international conference on extraction and knowledge management (EGC 2015), Abbey of Neumünster Cultural Exchange Center (ルクセンブルク, ルクセンブルク), 2015. 01. 28

- ② Hirotaka Oba, Masaharu Hirota, Richard Chbeir, Hiroshi Ishikawa, Shohei Yokoyama, “Towards Better Land Cover Classification Using Geo-Tagged Photographs”, IEEE International Symposium on Multimedia (ISM2014), The Splendor Hotel Taichung (台湾, 台中), 2014. 12. 10

- ③ 後藤 康路, 山田 竜平, 山本 幸生, 石川 博, 横山 昌平, “階層型 SOM に基づいた大規模月地震波形の可視化システム”, 第 6 回データ工学と情報マネジメントに関するフォーラム (DEIM2014), 淡路夢舞台&ウェスティン淡路(兵庫県, 淡路市), 2014. 3. 3

- ④ 白井 元浩, 廣田 雅春, 石川 博, 横山 昌平, “ジオタグ付き写真を用いたホットスポットの分類と関連の抽出”, 第 6 回データ工学と情報マネジメントに関するフォーラム (DEIM2014), 淡路夢舞台&ウェスティン淡路(兵庫県淡路市), 2014. 3. 4

- ⑤ Masaki Omori, Masaharu Hirota, Hiroshi Ishikawa, Shohei Yokoyama, “Can Geo-tags on Flickr Draw Coastlines?”, International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2014), Renaissance Dallas Hotel (Dallas, Texas, USA), 2014. 11. 05

- ⑥ Motohiro Shirai, Masaharu Hirota, Hiroshi Ishikawa, Shohei Yokoyama, “A method of area of interest and shooting spot detection using geo-tagged photographs”, ACM SIGSPATIAL Workshop on Computational Models of Place 2013 at ACM SIGSPATIAL GIS 2013, Orlando USA, 2013. 11. 5

- ⑦ Yasumichi Goto, Ryuhei Yamada, Yukio Yamamoto, Shohei Yokoyama, Hiroshi Ishikawa, “SOM-based Visualization for Classifying Large-scale Sensing Data of Moonquakes”, 4th International Workshop on Streaming Media Delivery and Management Systems, Compiègne France, 2013. 10. 30

- ⑧ 後藤 康路, 山田 竜平, 山本 幸生, 横山 昌平, 石川 博, “波形の類似性を考慮した大規模月地震データの可視化システムの実装”, 日本地球惑星科学連合大会, セッション P-PS23 月の科学と探査, 幕張メッセ(千葉県千葉市), 2013. 05. 23.

6. 研究組織

(1) 研究代表者

横山 昌平 (YOKOYAMA Shohei)
静岡大学・情報学研究科・講師
研究者番号: 20443236

(2) 研究分担者

石川 博 (ISHIKAWA Hiroshi)
首都大学東京・システムデザイン学部・教授
研究者番号: 60326014

(3) 連携研究者

山本 幸生 (YAMAMOTO Yukio)
独立行政法人宇宙航空研究開発機構・宇宙科学研究所・助教
研究者番号: 50425790

山田 竜平 (YAMADA Ryuhei)
独立行政法人国立天文台・RISE 月惑星探査検討室, 研究員
研究者番号: 60647379