

地球科学におけるデータ駆動型解析

桑谷 立^{*,**}

Data-driven Analysis in Earth Science

Tatsu Kuwatani *

*海洋研究開発機構地球内部物質循環研究分野 Department of Solid Earth Geochemistry, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka 237-0061, Japan. E-mail: kuwatani@jamstec.go.jp

**科学技術振興機構さきがけ PRESTO, Japan Science and Technology Agency (JST), 4-1-8 Honcho, Kawaguchi 332-0012, Japan.

キーワード : データ駆動科学, 逆問題, 機械学習, ベイズ推論, スパースモデリング

Key words : Data-driven science, Inverse problem, Machine learning, Bayesian estimation, Sparse modeling

(要旨) 地球科学では, 様々な観測・計測データをもとに地球の過去および現在の物理・化学的過程を明らかにすることが根本的命題とされる. 近年の観測・計測技術の発展は, これまでは得られなかったような大量もしくは新たな種類のデータ取得を可能にし, 地球システム・現象の理解に著しく貢献している. しかし, 精度・解像度・計測範囲などの限界の存在や, 得られるデータの種類に関する制限により, 我々が本当に知りたい物理量やプロセスを直接的に抽出することは未だに困難である. 一方, 近年の計算機技術の発達と数理・情報科学の進展は, 取得データから最大限に情報を抽出するための強力なデータ科学手法群を生み出し続けており, 機械学習や人工知能は産業界や実社会などで既に広く活用されている. これらはデータ駆動型解析として, 自然科学や工学分野においても応用が進んでおり, それぞれの分野に関する課題解決に貢献しつつある.

このような状況を背景に生まれた学問分野がデータ駆動科学である. データ駆動科学は, データの背後に潜む構造やプロセスの抽出に関して, 特定の対象や分野に依存しない普遍的な数理・情報科学的方法論の構築を目指している(Igarashi et al 2016). 2013 年度から, 文科省科研費補助金新学術領域研究「スパースモデリングの深化と高次元データ駆動科学の創成」

(略称: 疎性モデリング, 領域代表者: 岡田真人, <http://sparse-modeling.jp/>) と題した研究プロジェクトが開始し, 地球・惑星科学を含む生命科学・地学を中心とする幅広い自然科学分野において, 最先端のデータ解析手法を応用する研究が活発に行われている. 異なる分野が多数協働することで, ある分野で用いられている数理科学的手法が, 別分野にスムーズに伝搬され, 従来, 数理科学的な解決が困難だった問題に対しても, 新たな解決方法が見つかるようなケースも多数生まれてきている. 2016 年度からは, 科学技術振興機構の戦略的創造研究推進事業 CREST/さきがけ複合領域「計測技術と高度情報処理の融合によるインテリジェント計測・解析手法の開発と応用」(略称: 情報計測, 研究総括: 雨宮慶幸/北川源一郎, <https://www.jst.go.jp/kisoken/>) が発足し, 情報科学・データ解析手法により, 物理的・機械的な性能が規定する計測の限界突破を目的にした情報と計測の融合研究が進展しつつある.

冒頭にも述べたように, 地球科学分野では複雑なシステムに関する挙動と構造を常に不十分な観測・計測データから解明するという難解な逆問題に取り組まなければならない. そのため, 地球科学はデータを最大限に活用して情報抽出を行うデータ駆動型解析の恰好の適用対象といえよう. 我々の研究グループでは新学術領域「疎性モデリング」の発足に際して, 下記の参考文献に示した通り, 地球物理学・地球化学・地質学を含む幅広いアプローチの様々な対象に対して, データ駆動型解析を適用してきた. 例えば, 地球物理学においては, ベイズ推論的画像解析手法であるマルコフ確率場モデル(Ohno et al. 2012; 桑谷ほか 2014a)を地震波トモグラフィや地球内部の媒質評価に用いた研究(Kuwatani et al. 2014a,b)や, スパースモデリングの一種であるL1正則化法をGPS観測データからプレート境界のゆっくり滑りの空間分布を推定する問題に適用している(Nakata et al. 2016, 2017; 桑谷ほか 2016). 地球化学分野においては, 機械学習の代表的手法であるサポートベクトルマシン(SVM)をスパースモデリングの一種である全探索法と組み合わせて, 2011 年東北沖津波堆積物の地球化学的判別に適用した研究(Kuwatani et al. 2014c; 駒井ほか 2016)や, 主成分分析・因子分析・スパース重回帰分析などを津波堆積物や表層土壌堆積物の地球化学特性の評価に適用した研究などがある(桑谷ほか 2014b; 中村ほか 2016; Nakamura et al. 2016, 2017). また, 固体地球に関する地球化学分野においても, 独立成分分析・クラスタ分析などを火成岩の全岩化学組成や温泉水溶液組成に適用するなど多方面から研究を進めつつある(Iwamori et al 2017, 桑谷&岩森 印刷中).

地質学分野におけるデータ駆動型解析の先駆的な例としては, 変成岩中の鉱物組成累帯構造を用いた温度圧力(P - T)履歴推定に対してマルコフ確率場モデルを適用した研究がある(Kuwatani et al 2012). 最近では, 同様の問題に対して, 鉱物粒子の成長則を導入することで, データ同化の枠組みにより温度圧力時間(P - T - t)履歴の推定を可能にしている. また, 岩石形

成のダイナミクスを逆問題として抽出する研究例としては、岩石-水相互作用に関する室内実験系の時系列化学分析データから反応係数などを推定するデータ同化手法を開発した(Omori et al 2016). 今後は、より複雑で多様な岩石組織を対象にした岩石形成ダイナミクスの抽出法の開発を進めていく(Okamoto et al 2015; 岡本&桑谷 2017). 上記のほかにも、構造地質学の問題に情報量基準によるモデル選択法を適用した研究(Matsumura et al 2017)や、変成作用中の物質移動量の推定に非負値行列分解を適用した研究(Yoshida et al 2017)など、多数のデータ駆動型解析を活用した研究を進行中である。

本講演では、はじめに、データ駆動型解析のキーテクノロジーであるスパースモデリング(高次元データから少数の本質的な説明変数を自動抽出する数理的方法論)とベイズ推論(逆解析に順モデルと先験的知識を有効に導入する確率論的アプローチ)について地球科学分野以外の適用例も挙げながら解説する。その後、実際にデータ駆動型解析を地球科学分野に応用した例について、現在進行中の研究も含めて幾つか紹介する。最後に、情報地質学の発展にデータ駆動科学が果たすべき役割とともに、今後の展望について議論する。

参考文献

- Igarashi, Y, Nagata, K, Kuwatani, T, Omori, T, Nakanishi-Ohno, Y, Okada, M (2016) Three levels of data-driven science. *Journal of Physics: Conference Series*, 699, 012001:1-012001:13. doi:10.1088/1742-6596/699/1/012001
- Iwamori, H, Yoshida, K, Nakamura, H, Kuwatani, T, Hamada, M, Haraguchi, S, Ueki, K (2017) Classification of geochemical data based on multivariate statistical analyses: Complementary roles of cluster, principal component, and independent component analyses. *Geochemistry, Geophysics, Geosystems*, 18(3), 994-1012
- 駒井武, 桑谷立, 中村謙吾, 土屋範芳 (2016) スパースモデリングによる津波堆積物の判別-津波到達推定への応用-. *電子情報通信学会会誌*, 99(5), 418-423
- 桑谷立, 岩森光 (印刷中) ビッグデータ解析, 「図説 固体地球の事典」朝倉書店.
- Kuwatani, T, Nagata, K, Okada, M, Toriumi, M (2012) Precise estimation of Pressure - Temperature paths from zoned minerals using Markov random field modeling: theory and synthetic inversion. *Contributions to Mineralogy and Petrology*, 163(3), 547-562. doi.org/10.1007/s00410-011-0687-3
- Kuwatani, T, Nagata, K, Okada, M, Toriumi, M (2014a) Markov-random-field modeling for linear seismic tomography. *Physical Review E*, 90, 042137:1-042137:7. doi:10.1103/PhysRevE.90.042137
- Kuwatani, T, Nagata, K, Okada, M, Toriumi, M (2014b) Markov random field modeling for mapping geofluid distributions from seismic velocity structures. *Earth, Planets and Space*, 66, 5:1-5:9. doi.org/10.1186/1880-5981-66-5
- Kuwatani, T, Nagata, K, Okada, M, Watanabe, T, Ogawa, Y, Komai, T, Tsuchiya, N (2014c) Machine-learning techniques for geochemical discrimination of 2011 Tohoku tsunami deposits. *Scientific Reports*, 4, 7077:1-7077:6. doi:10.1038/srep07077
- 桑谷立, 永田賢二, 岡田真人(2014a) マルコフランダムフィールドモデルによる計測画像データからの潜在構造抽出. *映像情報メディア学会誌*, 68(12), 892-896
- 桑谷立, 中村謙吾, 渡邊隆弘, 小川泰正, 駒井武 (2014b) 主成分分析を用いた次元圧縮に基づく東北地方太平洋沖地震による津波堆積物の地球化学的特性評価. *地学雑誌*, 123(6), 923-935. doi.org/10.5026/jgeography.123.923
- 桑谷立, 中田令子, 岡田真人, 堀高峰(2016) スパースモデリングの地球物理学への応用. *電子情報通信学会会誌*, 99(5), 406-410
- Matsumura, T, Kuwatani, T, Masuda, T (2017) Statistical model selection between elastic and Newtonian viscous matrix models for the microboudin palaeopiezometer. *Earth, Planets and Space*, 69, 83:1-83:12. doi:10.1186/s40623-017-0669-4
- Nakamura, K, Kuwatani, T, Kawabe, Y, Komai, T (2016) Extraction of heavy metals characteristics of the 2011 Tohoku tsunami deposits using multiple classification analysis. *Chemosphere*, 144, 1241-1248. doi:10.1016/j.chemosphere.2015.09.078
- 中村謙吾, 佐藤海里, 川辺能成, 桑谷立, 駒井武(2016) 名取川水系の重金属類の環境評価に関する調査及び数理統計解析. *Journal of MMIJ*, 132(1), 22-30. doi:10.2473/journalofmmij.132.22
- Nakamura, K, Yasutaka, T, Kuwatani, T, Komai, T (2017) Development of a predictive model for lead, cadmium and fluorine soil/water partition coefficients using sparse multiple linear regression analysis. *Chemosphere*, 186, 501-509. doi:10.1016/j.chemosphere.2017.07.131
- Nakata, R, Kuwatani, T, Okada, M, Hori, T (2016) Geodetic inversion for spatial distribution of slip under smoothness, discontinuity, and sparsity constraints. *Earth, Planets and Space*, 68, 20:1-20:10
- Nakata, R, Hino, H, Kuwatani, T, Yoshioka, S, Okada, M, Hori, T (2017) Discontinuous boundaries of slow slip events beneath the Bungo Channel, southwest Japan. *Scientific Reports*, 7, 6129:1-6129:7. doi:10.1038/s41598-017-06185-0
- Ohno, Y, Nagata, K, Kuwatani, T, Shouno, H, Okada, M (2012) Deterministic algorithm for nonlinear Markov random field model. *Journal of Physical Society of Japan*, 81, 064006:1-064006:6. doi.org/10.1143/JPSJ.81.064006
- Okamoto, A, Kuwatani, T, Omori, T, Hukushima, K (2015) Free-energy landscape and nucleation pathway of polymorphic minerals from solution in a Potts lattice-gas model. *Physical Review E*, 92, 042130:1-042130:9. doi.org/10.1103/PhysRevE.92.042130
- 岡本敦, 桑谷立, (2017) 変成岩組織と鉱物組成帯構造からの情報抽出: フォワードモデルと逆解析. *地質学雑誌*, doi:10.5575/geosoc.2017.0034
- Omori, T, Kuwatani, T, Okamoto, A, Hukushima, K (2016) Bayesian inversion analysis of nonlinear dynamics in heterogeneous reactions. *Physical Review E*, 94, 033305:1-033305:11. doi:10.1103/PhysRevE.94.033305
- Wakasugi, K, Kuwatani, T, Nagata, K, Asoh, H, Okada, M (2014) Verification of effectiveness of a probabilistic algorithm for latent structure extraction using an associative memory model. *Journal of Physical Society of Japan*, 83, 104801:1-104801:8. doi.org/10.7566/JPSJ.83.104801
- Yoshida, K, Kuwatani, T, Hirajima, T, Iwamori, H, Akaho, S (2017) Progressive evolution of whole-rock composition during metamorphism revealed by multivariate statistical analyses. *Journal of Metamorphic Geology*, in press. doi:10.1111/jmg.12282