

R011-06

C 会場 : 11/24 PM1 (13:45-15:45)

15:15~15:30:00

太陽地球科学データへのシンボリック回帰の適用

#能勢 正仁¹⁾, 小山 聡¹⁾, 古賀 亮一¹⁾

(¹ 名古屋市立大学データサイエンス研究科)

Application of Symbolic Regression to Solar-Terrestrial Physics Data

#Masahito Nose¹⁾, Satoshi Oyama¹⁾, Ryoichi KOGA¹⁾

(¹ Graduate School of Data Science, Nagoya City University)

In solar-terrestrial physics, advances in observational instrument performance, multi-point observations, high temporal resolution, and increased storage capacity have led to the continuous production of vast datasets that are difficult to process by human effort alone. Traditionally, researchers have plotted observational data and applied statistical analyses to elucidate the properties of various electromagnetic phenomena in space, identify controlling parameters, and formulate relationships among physical quantities. However, with the exponential growth of data volume, it has become increasingly difficult to handle this information purely through manual analysis.

In information science and deep learning, a method known as "symbolic regression" has been developed to automatically derive mathematical models from data without incorporating human biases, relying on domain-specific knowledge, or pre-specifying functional forms. Proposed approaches include genetic algorithms, discovery of modular structure via neural networks, and hybrid frameworks that combine symbolic-regression and inference modules. Nevertheless, there are very few reports of applying these state-of-the-art techniques to real, large-scale observational datasets.

In this study, we focus on solar wind-magnetosphere coupling functions formulated in solar-terrestrial physics and investigate how these expressions—and related relationships among physical variables—are re-evaluated using symbolic regression. Among the many methods proposed (e.g., Eureqa, PySR, AI Feynman, AI-Descartes, and PhySO [Φ -SO]), we adopt PhySO, which offers publicly available code and demonstrated robustness to noise. On synthetic test data generated from theoretical equations, symbolic regression recovers the underlying formulae with high accuracy. Even when Gaussian artificial noise is added, and when working with OMNI data that include rounding errors, the estimates remain comparatively robust.

Because solar-terrestrial physics datasets almost invariably contain noise, symbolic regression may provide an effective tool for uncovering hidden governing equations. In this talk, drawing on concrete application examples, we discuss the potential of symbolic-regression methods to advance research in solar-terrestrial physics.

太陽地球系科学においては、観測器の性能向上・多点観測・高時間分解能・記憶媒体の大容量化が進み、人の能力だけでは処理することが難しいような大量のデータが連続して生み出されている。従来は、研究者が観測データをグラフ化したり、統計的処理を行ったりして、宇宙空間で生起している様々な電磁気現象の性質を解明し、それらをコントロールするパラメータを見出して物理量間の定式化を行ってきた。しかしながら、現在では、データ量が爆発的に増加し、人の能力だけでは処理することが難しくなっている。情報科学・深層学習の分野では、巨大なデータに対して、人間の先入観や分野の専門知識を含めたり、関数の形式を事前に指定したりせずに、データから数式モデルを自動的に生成する「シンボリック回帰」と呼ばれる手法が開発されてきている。具体的には、遺伝的アルゴリズム、ニューラルネットワークによるモジュラリティ性の発見、記号回帰モジュールと推論モジュールを組みわせる方法などが提案されている。ただ、こうした最先端のシンボリック回帰手法が実際の大規模観測データに適用されたという報告はほとんどない。

そこで、これまで太陽地球系科学分野で定式化が試みられてきた太陽風-磁気圏結合関数に焦点を当て、この式やそれに関係する物理量を導く式がシンボリック回帰によってどのように再評価されるかを調査した。シンボリック回帰の方法としては、Eureqa, PySR, AI Feynman, AI Descartes, Φ -SO など様々なものが提案されているが、コードが公開されており、ノイズに対する頑健性が高い Φ -SOを利用した。理論式から作成したテストデータに対しては、シンボリック回帰による物理式の推定は非常に優秀な成績を示した。正規分布の人工ノイズを加えたテストデータや、丸め誤差が含まれるOMNIデータについても、比較的良い推定結果が得られた。太陽地球系物理学のデータにはノイズが含まれることがほとんどであるが、そこから隠れた支配方程式を見出すのにシンボリック回帰は進展させる有効なツールになるかもしれない。講演では、実際の適用例を元に、シンボリック回帰手法の太陽地球系科学分野への応用可能性について議論する。