

R011-13

C会場 : 11/25 PM1 (13:15-15:15)

13:50~14:05

## シンボリック回帰の太陽地球系科学データへの適用による新たな支配方程式の発見

#能勢 正仁<sup>1)</sup>, 小山 聡<sup>1)</sup>, 古賀 亮一<sup>1)</sup>

<sup>1)</sup> 名市大・DS 学部

## Discovery of new governing equations through the application of symbolic regression to solar-terrestrial physics data

#Masahito Nose<sup>1)</sup>, Satoshi Oyama<sup>1)</sup>, Koga Ryoichi<sup>1)</sup>

<sup>1)</sup> School of Data Science, Nagoya City University

In solar-terrestrial physics, advancements in observational instrument performance, multi-point observations, high temporal resolution, and the large capacity of storage media have led to the continuous generation of vast amounts of data that are challenging to process solely by human capabilities. Traditionally, researchers have analyzed observational data through plotting and statistical processing to elucidate the properties of various electromagnetic phenomena occurring in space, identified the controlling parameters, and formulated the relationships between physical quantities. However, with the exponential increase in data volume, it has become increasingly difficult to handle this information purely with human effort. In the fields of information science and deep learning, a method called "symbolic regression" has been developed to automatically generate mathematical models from data without incorporating human biases, domain-specific knowledge, or pre-specifying the functional forms. Specifically, approaches such as genetic algorithms, discovery of modularity through neural networks, and combination of symbolic regression modules with inference modules have been proposed. However, there are very few reports on the application of these cutting-edge symbolic regression techniques to actual large-scale observational data.

Therefore, we focused on the solar wind-magnetosphere coupling functions, which have been the subject of formulation attempts in the field of solar-terrestrial science, and investigated how symbolic regression could re-evaluate these functions and the equations that derive related physical quantities. Various methods of symbolic regression, such as Eureqa, PySR, AI Descartes, and  $\Phi$ -SO, have been proposed. For this study, we utilized AI Feynman, which has publicly available code and relatively comprehensive documentation. When applied to test data generated from theoretical equations, the symbolic regression method demonstrated excellent performance in estimating physical equations. When artificial noise with a normal distribution was added to the test data, relatively good estimation results were obtained as long as the standard deviation of the noise was less than a few percent of the standard deviation of the test data. Since most data in solar-terrestrial physics contain some level of noise, symbolic regression could be an effective tool for advancing research when the noise is minimal. In the presentation, we will show actual application examples and discuss the applicability of symbolic regression methods to the field of 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 Descartes,  $\Phi$ -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、説明文章も比較的充実している AI Feynman 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。テストデータに正規分布の人工ノイズを加えたところ、ノイズの標準偏差が、テストデータの標準偏差の数 % 以下であれば、比較的良好な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り、それが小さい場合にはシンボリック回帰は研究を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。