

宇宙の大規模構造の複雑な統計パターンを高速予言する 人工知能 (AI) ツールを開発 —宇宙ビッグデータの AI 分析に向けて—

概要

京都大学基礎物理学研究所 西道啓博 特定准教授 (兼: 東京大学国際高等研究所カブリ数物連携宇宙研究機構 (Kavli IPMU) 客員科学研究員)、Kavli IPMU 高田昌広 主任研究者、弘前大学、国立天文台、名古屋大学の研究者からなる共同研究チームは、現在の宇宙で観測される銀河の大域的空間分布に見られる「網の目構造」の起源を調査するために、国立天文台のスーパーコンピュータ「アテルイ」及び「アテルイII」を用いて、大規模な宇宙の構造形成^(注1)シミュレーションを実行しました。ダークマター、ダークエネルギーなどの宇宙の組成^(注2)、またインフレーションモデルが予言する宇宙の初期条件に関するパラメータが構成する多次元空間から 101 個の代表点を選び出し、これら全てを網羅するシミュレーション群を実行することで、大規模なデータベースを構築しました。さらに、このデータベースを分析する AI フレームワーク「ダークエミュレータ」を開発し、「任意の」宇宙モデルにおける宇宙の大規模構造の観測量を正確かつ高速に計算することに成功しました。ダークエミュレータを用いることで、スーパーコンピュータでは数日かかる理論予言を、元々のシミュレーションの結果と遜色のない精度を保ちながらもノートパソコンで数秒以内に計算することが可能になります。これは計算コストをおおよそ 1 億分の 1 に低減したことになり、実観測データから宇宙の根源的な情報を引き出す操作を飛躍的に高速化することを可能としました。

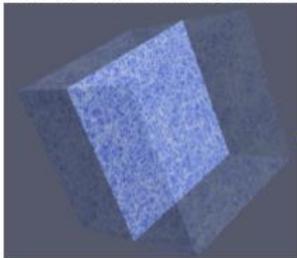
本研究成果は、2019 年 10 月 8 日に米国の天体物理学専門誌「The Astrophysical Journal」にオンライン掲載されました。

スーパーコンピュータシミュレーション
(約 100 億個の質量素片の宇宙年齢にわたる時間
進化を計算, 1 ランにつき 2 日)



国立天文台
スーパーコンピュータ
「アテルイII」

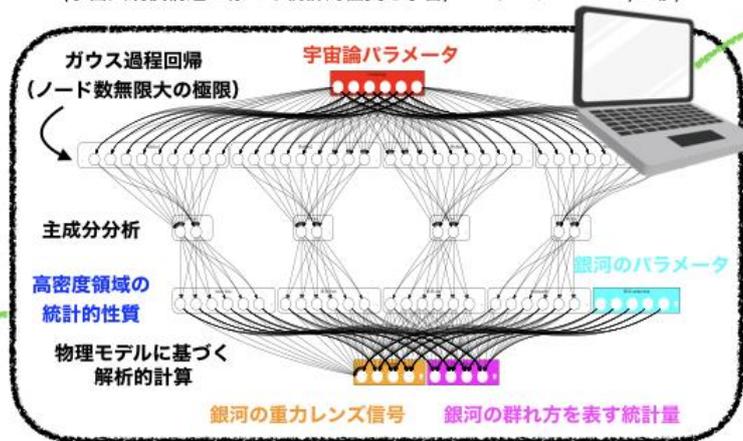
コンピュータ上に創られた仮想宇宙



学習

「ダークエミュレータ」ネットワーク

(宇宙大規模構造の様々な統計的性質を予言, ノートパソコンで 2, 3 秒)



予言



実観測データの高速かつ
正確な分析が可能に!

1. 背景

物理学の法則に基づく基礎理論と、近年の観測技術の目覚ましい進歩に支えられ、宇宙の基本的な枠組みを観測的に実証する「観測的宇宙論」は科学者が扱うことのできる最大スケールにおける実証科学として成熟してきました。特に、James Peebles 氏に贈られた 2019 年ノーベル物理学賞の主な研究対象であるビッグバン宇宙の残光「宇宙マイクロ波背景放射 (CMB)^{注3)}」に対する理論予言とその観測的実証は、宇宙論パラメータと呼ばれるわずか 6 つの基本的なパラメータのみで完全に記述される宇宙の「標準モデル」の確立へという形で実を結びました。しかし、この宇宙標準モデルの枠内では、ダーク成分と呼ばれるダークマター・ダークエネルギーといった素粒子標準理論を越えた未知の要素が宇宙の大半を占めており、このダーク成分の性質の理解なしには真の意味で宇宙の基本的な成り立ちを解明したことにはなりません。

CMB と並び、近年重要性が再認識されているのが「宇宙の大規模構造」です。これは、多数の銀河が密集した銀河団や超銀河団、それらを結節点として、空間上に糸のように張り巡らされたフィラメント、そして、広い領域に渡ってほとんど銀河が存在しないポイドなど、近傍宇宙に存在する銀河が織り成す網の目構造のパターンのことを指します(図 1 参照)。CMB が宇宙誕生後間もない原始の宇宙の様子を映し出している一方で、大規模構造はその後現在に至るまで複雑な進化を経た後の様子を描いています。従って、これを詳細に観測し、分析することで、宇宙の進化に大きな影響を及ぼす宇宙のダーク成分の謎に迫ることができると期待されています。実際、日本が主導するすばる望遠鏡を用いた大規模観測プロジェクト SuMIRe (Subaru Measurement of Images and Redshifts) をはじめ、世界中で大規模観測計画が稼働中または計画中です。

これらの観測から得られたデータを分析し、宇宙論的に有用な知見を引き出すためには、物理理論に基づいて、宇宙の構造進化に対して正確な予言を与える必要があります。スーパーコンピュータを用いた数値計算はこの目的にとって極めて有効であることが知られており、計算資源の増大に伴って高精度かつ大規模な計算が可能となってきました。しかし、観測データを分析して、宇宙のダーク成分の量や性質を調査するためには、異なるパラメータを持つモデルを多数用意し、それぞれについて理論予言を与え、観測データを最もよく再現するものを探し出す必要があります。特に、多次元の宇宙論パラメータ空間(標準的なもので 6 次元空間)においては数十万から 100 万にも及ぶパラメータの組み合わせに対して精巧な理論計算を行う必要があります。しかし、現在利用可能な最大の計算資源を以てしても、これほどの数のシミュレーションを実行することはできませんでした。

2. 研究手法・成果

本研究グループは、AI の一種である「機械学習」を利用する事でこの問題に取り組みました。脳の機能に似せて作られた「人工ニューラルネットワーク」に代表される機械学習は、入力と出力との未知の対応関係を複雑に組み立てられた神経回路のような構造を使って学び取る仕組みで、適切に「学習」させることで新たな予言ができるようになります。宇宙のダーク成分の量や性質などを様々に変えて、スーパーコンピュータにより多数のヴァーチャル宇宙を生成し、その対応関係を学び取らせ、新たなシミュレーションを実行することなく高速に予言を行おうというのが今回の試みの核心部分です。こうして作られた、本物を実行することなく手軽に同等の効果が得られる仕組みは、しばしば「エミュレータ」と呼ばれます。

しかし、シミュレーションにより精密な宇宙モデルを生成するためにかかる計算コストが高いため、膨大な数の学習データを用意することができません。そこで、本研究グループは多次元宇宙論パラメータ空間から緻密な実験計画を立て、101 個という比較的少ない代表点を最適な形で抽出し、地球統計学などで比較的古くか

ら実績のある「ガウス過程回帰^(注4)」という手法を中核に据えた、多層から成る複合的なネットワーク（図2参照）により学習させました。この方法では、学び取りたいパターンの複雑さをネットワーク中枢にあるごく少数のパラメータのみがコントロールしているため、比較的少数の学習データでも十分性能を発揮します。また、ネットワーク内部では、シミュレーションだけからは完全に表現し切れない、銀河形成などの複雑な過程に関するバリエーションを、物理に基づく数理モデルで表現しているため、現実の宇宙の複雑な銀河のデータに十分柔軟に対応して予言を与えることができます。

「ダークエミュレータ」と名付けられた本研究グループの学習装置は、宇宙の大規模構造の理論予言に必要なあらゆる基本的な統計量を学習しました。この学習には、代表的な101組の宇宙論パラメータに対して、国立天文台のスーパーコンピュータ「アテルイ」および「アテルイII」を用いて約3年かけて実行された、計262試行ものシミュレーションデータから成る、総容量約300テラバイトの巨大データベースが用いられました。学習結果を詳細に検証した結果、ダークエミュレータは、すばる望遠鏡に搭載された超広視野主焦点カメラ（Hyper Suprime-Cam; HSC）が捉える弱重力レンズ効果^(注5)による信号や、スローン・デジタル・スカイ・サーベイ（SDSS）による銀河の3次元空間上の分布パターンの観測データを、概ね誤差2~3%程度の精度で予言できることが実証されました。ダークエミュレータは、与えられた宇宙論パラメータ及び銀河の複雑さを表すパラメータに対し、標準的なノートパソコンでも数秒以内に理論予言を行うことができます。これは計算コストをおおよそ1億分の1に低減したことになり、実観測データから宇宙の根源的な情報を引き出す操作を飛躍的に高速化することを可能としました。

3. 波及効果・今後の展望

宇宙論における海外の同様な取り組みでは、観測量に直接結びつく理論予言、特に、現実世界の銀河が持つ多様性を適切に取り扱えるだけの柔軟性が確保されておらず、更に、本研究グループが行ったような信頼性の高い誤差の評価がなされていませんでした。ダークエミュレータは、実際の観測データに初めて直接応用することができるAIツールと言えます。ダークエミュレータは、すばる望遠鏡の宇宙の大規模構造データを分析できるのはもちろん、2020年代中盤以降に控える次世代の究極的な宇宙論観測の時代における、AIを用いた宇宙ビッグデータ分析への着実な一歩と言えます。

4. 本研究プロジェクトについて

本研究は、科学技術振興機構 CREST (JPMHCR1414、代表：吉田直紀)、科学研究費補助金 新学術領域研究「なぜ宇宙は加速するのか？-徹底的究明と将来への挑戦-」総括班 (15H05887、代表：村山斉) 計画研究 (15H05892、代表：宮崎聡、15H05893、代表：高田昌広)、文部科学省ポスト「京」重点課題9「宇宙の基本法則と進化の解明」および計算基礎科学連携拠点 (JICFuS)、科学研究費補助金 若手研究 (B) (17K14273、代表：西道啓博)の援助を受けています。

<用語解説>

注1) 宇宙の構造形成：晴れ上がり後の宇宙の構造形成においては重力が主たる役割を果たす。宇宙の平均と比べて僅かに密度が高い領域は重力により周囲の物質をかき集め、逆に密度の低い領域からは物質は周囲に吸い出される。このような過程を通じて、初期宇宙の微小な揺らぎが徐々に増幅され、現在の宇宙に見られる豊かな構造が形成されたと考えられている。

注 2) 宇宙の組成：現在の宇宙の全エネルギーに対して異なる成分がそれぞれ占める割合。銀河の内部の星がバラバラにならず留まるために必要で、なおかつ宇宙の構造形成を説明するために不可欠な「ダークマター」、宇宙の加速膨張を引き起こしている未知のエネルギー源「ダークエネルギー」が合わせて約 95%を占めると考えられている。素粒子物理学により知られている通常の物質は約 5%に過ぎない。

注 3) 宇宙マイクロ波背景放射：ビッグバン直後の灼熱、高密度の宇宙では、光は頻繁に散乱され、真っ直ぐ進むことができない。やがて宇宙が膨張し、温度が冷えていくと、この状況が解消され、光はまっすぐに進むようになる（宇宙の晴れ上がり）。この光は温度約 2.7K のマイクロ波として観測され、光が届く方向に応じて僅かにだけ温度の高低が見られる。この方向に応じたパターンは宇宙晴れ上がり期の「宇宙論的揺らぎ」を表している。その精密測定を通じて現在の標準的宇宙モデルが導き出された。

注 4) ガウス過程回帰：地球統計学などの分野で広く使われてきた「ノンパラメトリックベイズ推定」と呼ばれる一種の計算手法。例えば、疎らに置かれた観測点において測定された気温などの気象データから、観測点の無い任意の場所における気象情報を予測する際などに使うことができる。ガウス過程はその空間パターンの複雑性などを調整する内部パラメータを含んでおり、これらを実際の観測データに最も合うように調整することで、一種の機械学習として機能する。今回の発表では 2 次元地表面の代わりに 6 次元宇宙論パラメータ空間が入力となるが、ガウス過程は次元が高くなっても高い性能を示すことが知られている。

注 5) 弱重力レンズ効果：アインシュタインの一般相対性理論が予言する効果で、伝搬中の光が周囲の構造が作る重力場によって曲げられるため、銀河の像が歪んで見えることを指す。個々の銀河に対する歪み方は微弱であるが、すばる望遠鏡 HSC などを使って大量の銀河の像を測定することで、この効果を統計的に抜き出すことができる。光の曲がり方から逆算することで、観測した銀河と我々との間にある構造の濃淡（揺らぎ）を調査することができる。

<研究者のコメント>

スーパーコンピュータを用いて 3 年かけて作り上げた巨大データベースが、手元の PC でほんの数秒で再現できるようになりました。データ科学の手法の大きな可能性に手応えを感じています。この成果を応用して、現代物理学最大の難問とされる「ダークエネルギー」などの宇宙の根源的謎に迫りたいです。また、本研究の手法は、多自由度を持ち計算コストが大きい自然科学・社会科学の諸問題に対して広く応用できると期待しています。

<論文タイトルと著者>

タイトル：Dark Quest. I. Fast and Accurate Emulation of Halo Clustering Statistics and Its Application to Galaxy Clustering (ダーククエスト I. ハロークラスタリング統計量の高速精密エミュレーションと銀河クラスタリングへの応用)

著者：Takahiro Nishimichi (1,2), Masahiro Takada (1), Ryuichi Takahashi (3), Ken Osato (4), Masato

Shirasaki (5), Taira Oogi (1), Hironao Miyatake (1,6,7), Masamune Oguri (1,4,8), Ryoma Murata (1,4), Yosuke Kobayashi (1,4), and Naoki Yoshida (1,4,8)

著者所属：

1 Kavli Institute for the Physics and Mathematics of the Universe (WPI), The University of Tokyo, Chiba, Japan

2 Center for Gravitational Physics, Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto, Japan

3 Faculty of Science and Technology, Hirosaki University, 3 Bunkyo-cho, Hirosaki, Aomori, Japan

4 Department of Physics, The University of Tokyo, Tokyo, Japan

5 Division of Theoretical Astronomy, National Astronomical Observatory of Japan, Mitaka, Tokyo, Japan

6 Institute for Advanced Research, Nagoya University, Nagoya, Japan

7 Division of Particle and Astrophysical Science, Graduate School of Science, Nagoya University, Nagoya, Japan

8 Research Center for the Early Universe, The University of Tokyo, Tokyo, Japan

雑誌名：The Astrophysical Journal

DOI： <https://doi.org/10.3847/1538-4357/ab3719> (2019年10月8日掲載)

*論文のアブストラクト (The Astrophysical Journal のページ)

<https://iopscience.iop.org/article/10.3847/1538-4357/ab3719>

*プレプリント (arXiv.org のウェブページ)

<https://arxiv.org/abs/1811.09504>

<参考図>

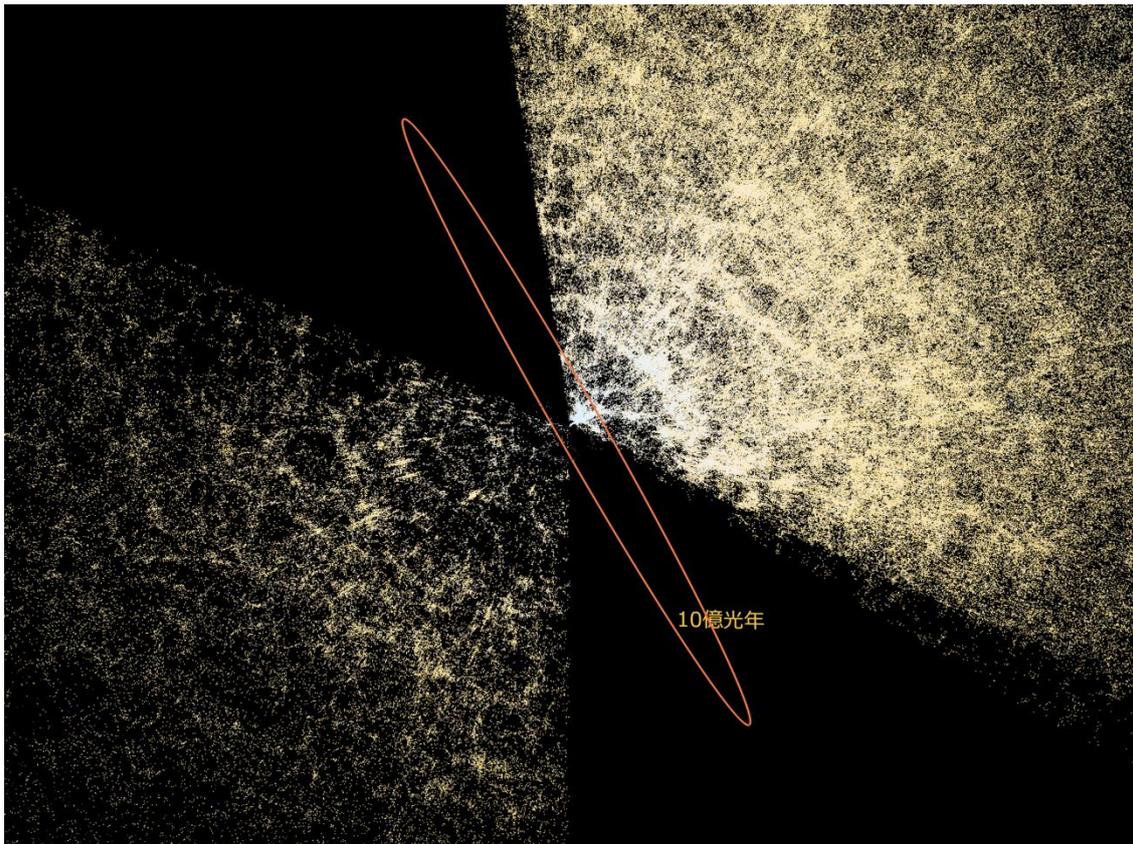


図 1

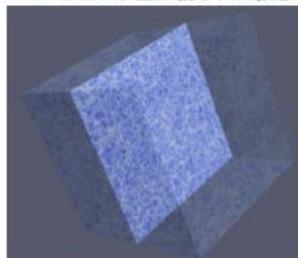
スローン・デジタル・スカイサーベイ (SDSS) が観測した、銀河が織り成す宇宙の大規模構造。黄色い点は銀河を、オレンジ色の円は半径 10 億光年の円を表す。中央に観測点である地球が存在し、扇形の観測領域内に存在する銀河の 3 次元空間分布図が描かれている。銀河が一様に存在するのではなく、網の目のような構造を成しながら分布している様子が見て取れる。(Credit: 加藤恒彦、ARC and SDSS、国立天文台 4 次元デジタル宇宙プロジェクト)

スーパーコンピュータシミュレーション
 (約100億個の質量素片の宇宙年齢にわたる時間
 進化を計算, 1ランにつき2日)



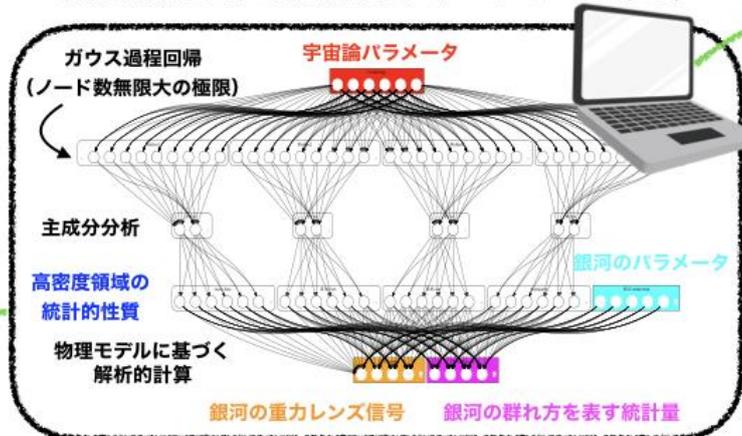
国立天文台
 スーパーコンピュータ
 「アテルイII」

コンピュータ上に創られた仮想宇宙



学習

「ダークエミュレータ」ネットワーク
 (宇宙大規模構造の様々な統計的性質を予言, ノートパソコンで2,3秒)



予言



実観測データの高速かつ
 正確な分析が可能に!

図2 ダークエミュレータの概略。左はスーパーコンピュータ「アテルイII」により計算した宇宙の大規模構造の一例。この計算は約100億の粒子を一辺が約49億光年にわたる領域に配置し、現在まで進化させたもので、「アテルイII」で800コアを同時に使用して約2日を要する。中図はこうして得られた仮想宇宙の大規模構造を学習する数値コード「ダークエミュレータ」の設計図。シミュレーション内に採用された宇宙の基本的枠組みを決める「宇宙論パラメータ」と、得られた大規模構造との対応関係を、複数の統計的手法と物理モデルとを組み合わせたハイブリッド学習機械により学習する。訓練されたダークエミュレータは、シミュレーションと遜色のない高精度で宇宙の大規模構造の複数の統計量を瞬時に予言する。これを利用すれば観測データを分析し、宇宙の基本的パラメータを引き出す分析にかかる計算コストを圧倒的に短縮できる。(Credit: YITP、国立天文台)