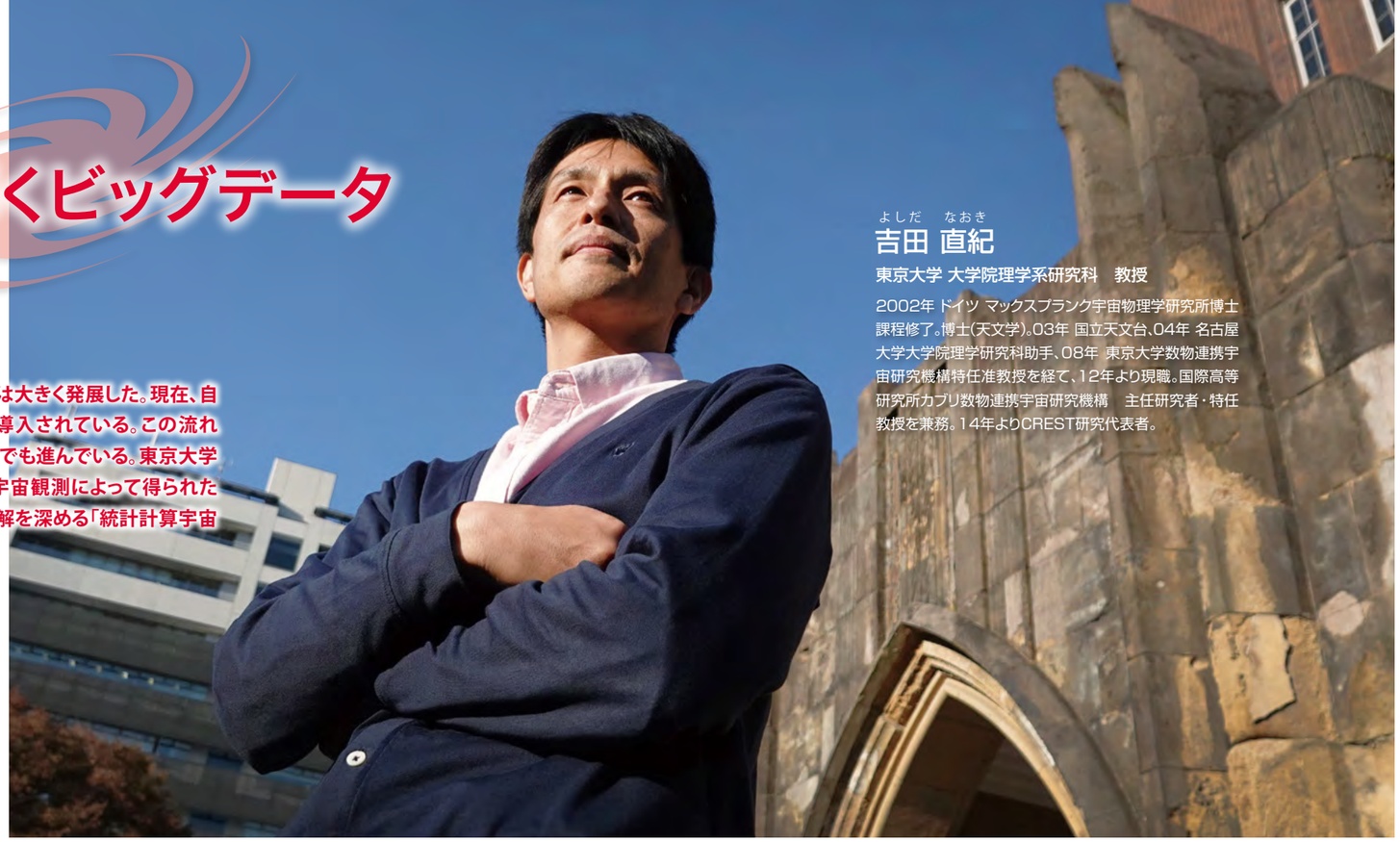
宇宙を読み解くビッグデータ



よしだ　なおき
吉田 直紀
東京大学 大学院理学系研究科　教授

2002年 ドイツ マックスプランク宇宙物理学研究所博士課程修了。博士(天文学)。03年 国立天文台、04年 名古屋大学大学院理学研究科助手、08年 東京大学数物連携宇宙研究機構特任准教授を経て、12年より現職。国際高等研究所カブリ数物連携宇宙研究機構　主任研究者・特任教授を兼務。14年よりCREST研究代表者。

この10年ほどの間に、人工知能（AI）の技術は大きく発展した。現在、自動運転、医療、金融などさまざまな分野で導入されている。この流れは、AIとあまり縁がなさそうな天文学の分野でも進んでいる。東京大学大学院理学系研究科の吉田直紀教授は、宇宙観測によって得られたビッグデータをAIによって分析し、宇宙の理解を深める「統計計算宇宙物理学」の創出に取り組んでいる。

変貌する天文学研究 観測データは膨大に

天文学という言葉から、一晩中、望遠鏡をのぞいて天体を観測する様子を思い浮かべる人も多いだろう。しかし、現代の天文学の現場では、そのような姿はほとんど見られない。

最先端の研究では、口径8メートル以上の大型望遠鏡を使うのが主流になっている。研究者が望遠鏡を直接のぞくのではなく、別室に設けられたコントロールルームで、たくさんの画像を撮影していく。

日本のすばる望遠鏡も8.2メートルの口径をもつ。ハワイの最高峰マウナケア山頂に設置された世界最高レベルの観測環境で、研究者が活躍している（図1）。

すばる望遠鏡の特徴は、さまざまな観測機器を取り付けられ、可視光から近赤外線まで幅広い観測ができることだ。特に、2012年8月から運用が始まった超広視野カメラのハイパー・シュプリーム・カム（HSC）では、満月10個分もの領域を一度に撮影できる。いわば、世界最大の「デジタルカメラ」だ。

ただし、HSCは視野が大きい分、データ量も大きくなる。市販されている一眼レフカメラで撮影した画像は大きいものでも5000万画素程度だが、HSCで撮影した画像は1枚で10億画素にもなる。夜空を一晩撮影するだけでも、得られる画像データは数百ギガ（10億）バイトだ。5年かけて取得する300夜分の観測データは、全部で25兆画素、1ペタ（1000兆）バイト分になるという。画像データには、およそ5億個の銀河、50万個のブラックホール、2万個の超新星が含まれる。

東京大学大学院理学系研究科の吉田直紀教授は、計算科学や統計学の手法を用いてこの膨大な天文データを解析し、宇宙の謎に迫ろうとしている。

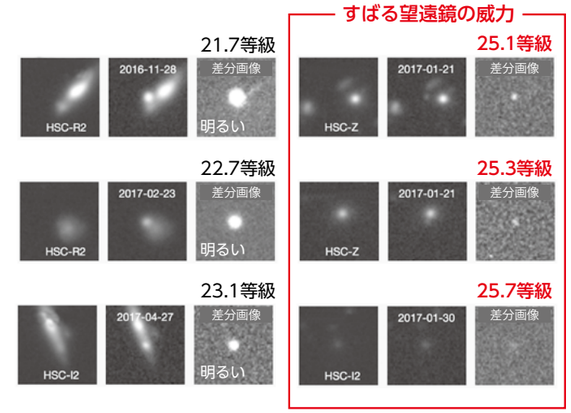
■図1　すばる望遠鏡
標高4200メートルのハワイ島マウナケア山頂にある大型光学赤外線望遠鏡「すばる」。光を集める鏡の有効口径は世界最大級の8.2メートルで、高い鏡面精度を誇る。鏡の形状を最適に保つ制御機構や、空気の乱れを押さえる新型ドーム、観測装置の自動交換システムなど、解像度や観測能率向上のための多くの工夫が施されている。© 国立天文台

コンピューターの目で 画像を素早く解析

1枚の画像データの中には、肉眼で見えるものだけでも数百個の天体が写っている。肉眼では判別できない暗い天体まで含めると、その数は数千個にまで膨れ上がるという。吉田さんは「1枚の画像データに写っている天体を種類や形態、検出法などに従ってリストにまとめ、カタログ化するだけでも一苦労ですし、その中から科学的に面白いものを発見するのは、人の力では難しくなっています」と説明する。

そこで、注目したのがAI技術だ。コンピューター技術の発展に伴い、近年、AIによる画像解析技術は急速に性能を上げている。HSCの膨大な観測データを解析し、宇宙について新しい知見を得るために考えたのが、「統計計算宇宙物理学」という新たな学問領域だ。

「天文学では、観測と理論の両面から研究されてきました。観測データも増えているのですが、理論研究の方も、スーパーコンピューターによるシミュレーションが盛んになっていて、たくさんのデータが出てきます。膨大な量のデータを数学的に効率よく比較するためにも、最先端のデータサイエンスの手法を取り入れたいと思ったのです」。



■図2　AIが複数日の撮影画像の差分から超新星を見分ける。試行錯誤の結果、判別精度の向上に成功した。HSCの高解像度画像により2等級以上暗い超新星（赤枠内）の解析が可能になったことも、超新星の発見に大きく貢献した。

チャレンジ精神を共有し それぞれの専門を生かす

2014年に開始したCRESTの研究には、東京大学の他に統計数理研究所、NTTコミュニケーション科学基礎研究所、筑波大学、慶應義塾大学のグループが参加している。HSCで取得した画像を解析するという大きな目的の下、天文学、宇宙物理学、計算科学分野の専門家が集まったものの、開始当初は手探り状態だったという。

「どのグループもそれぞれの専門分野で最先端の研究に取り組みたいという思いをもっていました。そのチャレンジ精神を共有しつつCRESTの研究テーマを設定するのに苦労しました。何度も集まって、それぞれの専門を生かして何ができるのかを話し合い、少しずつ研究のターゲットを絞っていきました」と振り返る。

そして、コミュニケーションを重ねるうちに、機械学習を「超新星爆発の探索」と「ダークマターの3次元地図作成」に利用する2つの手法が開発されていった。

画像の差分から 超新星爆発を発見

超新星は、太陽よりも質量の大きな重い恒星が、その一生を終えるときに起こす大きな爆発のことで、超新星爆発とも呼ばれる。宇宙全体では3秒に1回の割合で起こっているとされるが、実際には宇宙の限られた領域しか観測できなかったため、従来の観測では一晩に数個が見つかる程度であった。

超新星爆発が起きているかどうかは、数日分の画像を見比べればわかる。例えば、前の日に撮影した画像と見比べて、明るさが変化した天体が見つかれば、超新星爆発が生じた可能性がある。従来は人が異なる日に撮影されたデータを見比べて、1つ1つ探していたが、データ量が爆発的に増加ことにより、分析が追い付かなくなっていた。

吉田さんは、観測日の違う2つの画像の差分画像を作ることで、超新星爆発の候補を浮かび上がらせようと考えた。ここで活用したのがAIだ。観測画像や人工天体を含む2万4000個のデータから超新星爆発の画像の特徴を学んだAIが、データに含まれる多くのノイズの中から超新星爆発を起こした天体を発見するのだ（図2）。

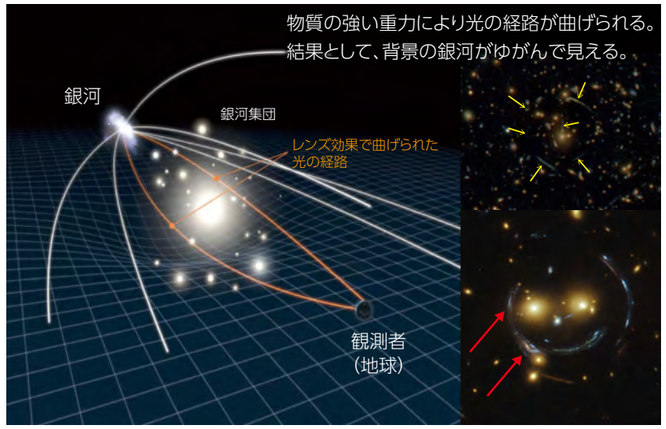
変化の特徴からIa型を選別 宇宙の歴史を知る手掛かりに

超新星爆発にはいくつかの種類がある。中でも「Ia型超新星爆発（Ia型）」はピーク時の明るさが一定で、観測されるIa型の見かけの明るさから地球からの距離を見積もることができる。2011年にノーベル物理学賞を受賞した宇宙の加速的膨張の発見は、Ia型の観測による成果だ。天体までの距離を測る「宇宙の物差し」として、宇宙が進化してきた様子を知り、未来を予測する上で重要な観測対象となっている。吉田さんは、機械学習によって発見した超新星爆発の中からIa型のみを選び出すことにも成功した。

「Ia型は、いったん、明るく光った後に、数日から10日ほどで急激に暗くなるという特徴をもっています。コンピューターは2万4000個の人工天体データから、そういったIa型の特徴を学びます。その結果、数日間隔で撮影した画像をコンピューターに読み込むと、Ia型かどうかを瞬時に判断してくれるようになったのです」。

AIの利用により、変動が見られた6万5000個の天体のうち、1824個が超新星爆発であることが突き止められた（図3）。さらに、これらを分類し、25個のIa型を確認した。いずれも地球から94億光年以上離れている。天体からの光が地球に届くまで時間がかかるため、私たちは過去の光を観察している。つまり、宇宙では遠くを見れば見るほど、過去の様子を知ることができるのだ。今回の観測で、遠くのIa型が多数発見されたので、この宇宙がこれまでどのように発展してきたのか、そして、今後どのようになってくのかが、よりはっきりしてくる。場合によっては、宇宙に対する理解を大きく変えてしまう可能性も秘めている。

■図3　膨大な画像の中から、一晩で数十個の超新星を発見した。人の目では不可能なペースだ。



■図4　重力レンズ効果の模式図（左）とすばる望遠鏡で撮影された重力レンズ効果（右）。観測者と銀河の間に大きな質量をもつ物体が存在すると、重力の影響を受けた周囲の空間がゆがめられ、光の経路が曲げられる。その結果、銀河が弓状に変形して観察される（矢印）。©左・NASA/ESA & L. Calçada、右上：東京大学／国立天文台／大栗真宗助教（東京大学）、右下：NASA／ESA

見えないダークマターの空間分布をあぶり出す

CRESTでのもう1つのターゲットが、「ダークマター」だ。正体がまだ確認されていない謎の物質だが、宇宙の物質の大部分を占め、この宇宙が現在の状態になるのに大きく関わっているとされる。人には見ることも、触ることもできないが、吉田さんは「重力レンズ効果」という現象を利用して、ダークマターの宇宙空間での分布を明らかにしようとしている。

宇宙では、大きな質量がある場所は空間がゆがみ、光が真っすぐに進めずに曲がってしまうという現象が起こる。これを重力レンズ効果という（図4）。それにより、遠くに位置する銀河の姿は地球に届くまでにゆがんでしまう。

遠くの銀河の形がゆがむということは、その銀河から地球までの間に大量の質量をもつ物質があるということを意味している。多くの場合、重力レンズ効果によって計算される質量は、望遠鏡で観測できる銀河の質量を上回る。つまり、計算された質量から、観測された銀河の質量を差し引いたものが、観測することのできないダークマターの質量ということになる。この考え方で、どの場所に、どのくらいの量のダークマターが存在するのかが徐々に明らかになってきた。

3次元地図で可視化 機械学習は試行錯誤

研究チームが開発したのは、HSC画像からダークマターの2次元分布を割り出し（図5）、それをさらに3次元地図として可視化する手法だ（図6）。HSCにより宇宙の広い範囲の画像を高解像度で得られる。この画像に写っている2000万個以上の銀河の形を解析し、そのゆがみ具合からダークマターの2次元分布を逆算する。重力レンズの効果による画像のゆがみは、観測点と銀河の中間にダークマターが存在するときに最大になる。これを利用して、銀河の距離ごとに解析することで、2次元分布から3次元画像を再構築できる。

一方、機械学習を用いた解析手法により、2次元の画像データから個々の銀河に付随するダークマターの量を正確に推定することにも成功した。コンピューターの学習に用いたのは、天文学で過去に得られた銀河の形とダークマターの質量などを関連付けたデータだ。実験では、35万枚の画像を学習させ、ダークマターの量を瞬時に計算できるまでになった。

吉田さんは「これまでは、1つ1つの画像に対して、ダークマターの量を手作業に近い形で見積もっていたのですが、AIの場合は、正しく学習させれば、すぐに答えを返せるようになります」と、その威力を語る。

ただし、ここまで学習させるには試行錯誤を繰り返した。人の場合は、画像を見ただけで、どこに注目すればいいのかがすぐにわかるが、コンピューターの場合はそうはいかない。いくつかのフィルターを使い、色分けをして、コンピューターが認識しやすい状態を探っていった。「人とは理解の仕方が違うので、もどかしいところもありましたが、『赤い波長の遠くにある銀河は関係があって、青い波長の近い銀河は関係がない』というように、どこに注目するのかを定義してあげると、正確な数値を計算するようになりました」。

ダークマターの3次元地図は、宇宙の構造や成り立ちに関する従来のモデルや理論を検証する重要な材料となる。可視光以外にも電波やX線データの利用や、機械学習手法の検討など改良を重ね、さらに精度を向上させる計画だ。

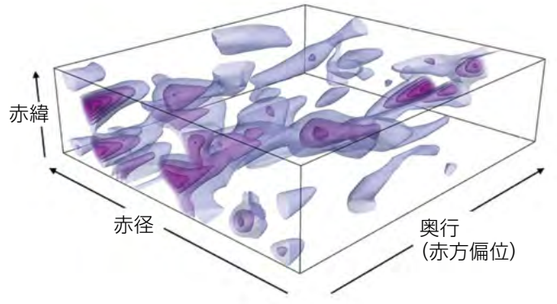
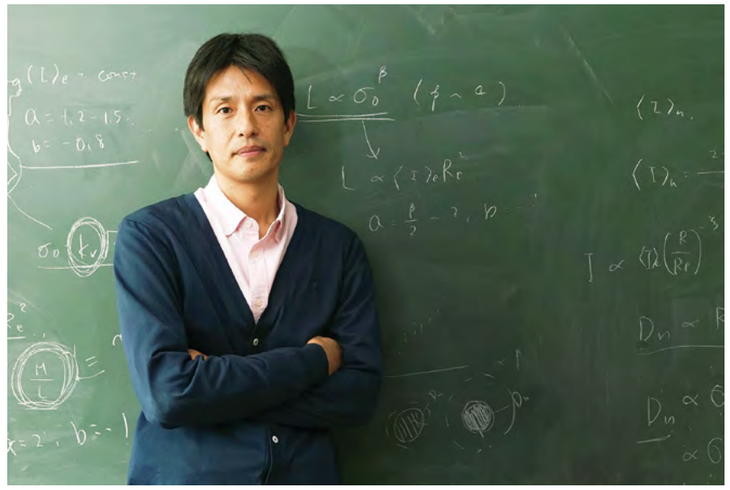
統計学との融合で 天文学を新たな領域へ

今後は、観測結果をシミュレーションの結果と直接照合するようにしていきたいと吉田さんは考えている。例えば、少しずつ違う理論に基づいたダークマターの分布のシミュレーション結果と、今回得られたダークマターの3次元地図を比べれば、どの理論が正しいかが高い精度で判定できる。

「CRESTで取り組んでいることは、新しい分野の開拓です。天文学はもうAIのサポートなしには研究できない状況になっています。今は開拓中の技術ですが、将来の天文学の発展につながればと考えています」と吉田さんは期待を膨らませる。

天文学の世界は、今や夜空の星を観測するだけでなく、膨大な観測データから宇宙がどのように広がってきたのかを精密に推定する時代となった。他分野の多様な研究領域と同様、この宇宙をより精密に理解するためにもAIの手助けが必要だ。吉田さんたちは、その基礎を築いている。

■図5　HSCで観測された天体画像の一部と、解析で得られたダークマターの2次元分布（等高線）。©国立天文台／HSC Project



■図6　ダークマターの3次元分布図。HSCの大集光力により、10億光年×2.5億光年の範囲、さらに奥行は約80光年という広大な天域での3次元分布が得られている。©東京大学／国立天文台／大栗真宗助教（東京大学）