

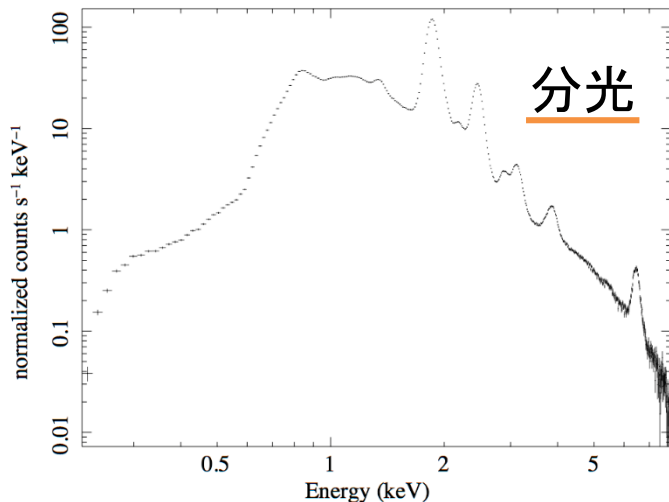
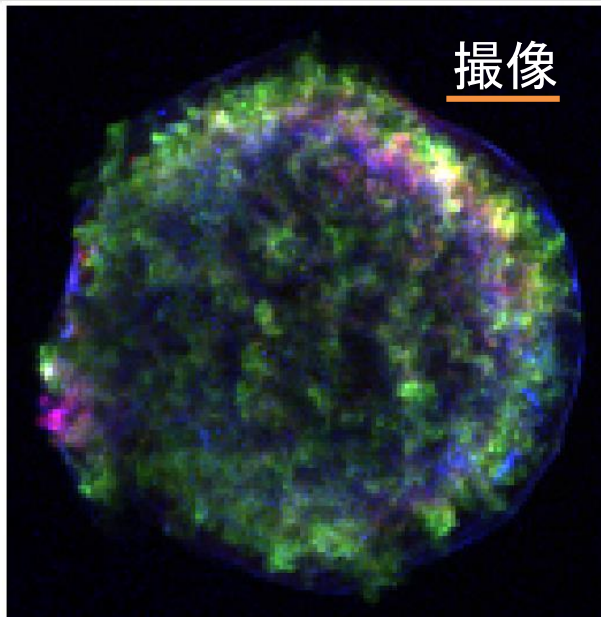
# X線天文解析における深層学習VAEを用いた特徴抽出

---

岩崎啓克、一戸悠人、内山泰伸(立教大)、山口弘悦(宇宙研)

宇宙科学情報解析シンポジウム

2019年2月15日



天文観測データは高次元データ  
(空間、時間、波長など)

X線観測は同時に撮像 & 分光  
空間と光子エネルギーの情報を持った  
複雑なデータが得られる

現状では、人の目で特徴を把握し、  
人の手で領域選択する

現在、豊富な観測データが利用可能

- Chandra (1999~)
- XMM Newton (1999~)
- Suzaku (2005~2015)

詳細な解析が十分進んでいない天体もある  
人が見落としてきた特徴もありうる

# 天文+機械学習 | モチベーション

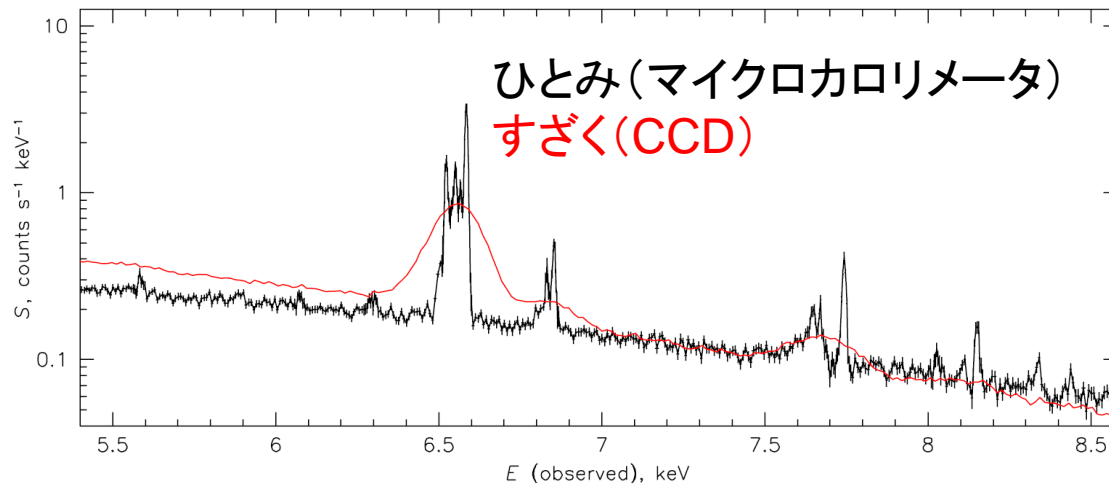
さらに、将来のX線観測機計画では、マイクロカロリメータによって  
波長分解能が飛躍的に向上する  
(150 eV→2.5 eV @6 keV × 約4000 pix, Athena)



人の手に負えないかもしれない。  
機械的な特徴発見・解析をしたい。



Athena 衛星 (ESA)



Hitomi Collaboration 2016

# 天文+機械学習 | モチベーション

天文観測データは高次元データ  
(空間、時間、波長など)

- 人の目で特徴を捉えるのは大変
- バイアス、見落としの可能性



複雑なデータから機械的・無バイアス  
に特徴を見つけたい。

将来の観測機計画では波長分解能  
が圧倒的に向上する

(150 eV → 2.5 eV @6 keV  
× 約4000 pix, Athena)



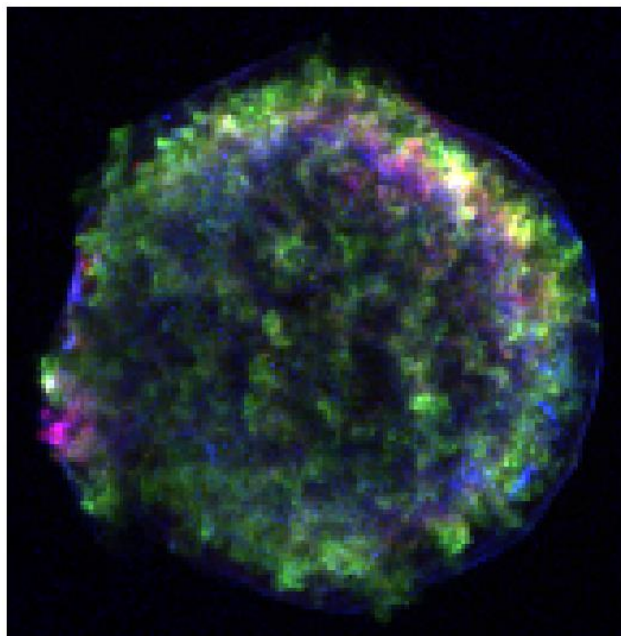
人の手に負えないかもしれない。  
機械的な特徴発見・解析をしたい。



## 機械学習を応用

手始めに、超新星残骸(SNR)のスペクトル情報  
から特徴的な空間構造を見つける手法を考案  
した。

ベンチマークとしてTycho's SNRでデモを行う。



# 深層学習

最近、天体観測やシミュレーションの解析に深層学習を応用する機運が高まっている。

例) 銀河の分類、スペクトル解析

## ◆ 深層学習

深層ニューラルネットを使ったモデルを学習させる。

多層のニューラルネットは非線形な特徴の表現が可能。

ビッグデータ、高次元データから特徴を抽出するのが得意。

天文のデータはまさしく高次元、ビッグデータ

## ■ 教師あり学習

データと正解ラベルがセットになっている。

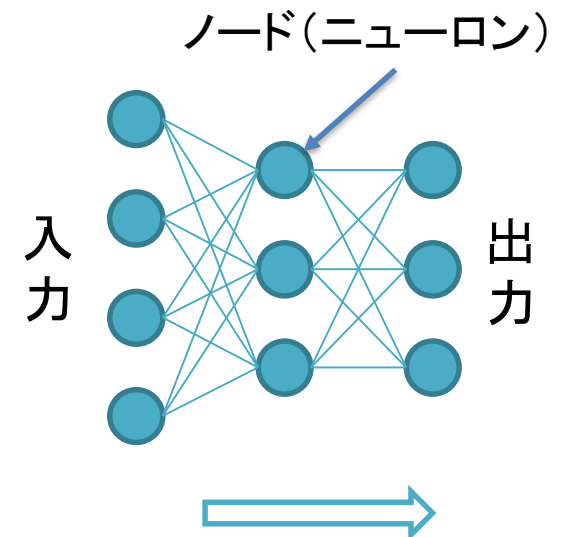
~数万個のデータ点への正解の付与は高コスト

## ■ 教師なし学習

データに正解ラベルがついていない。

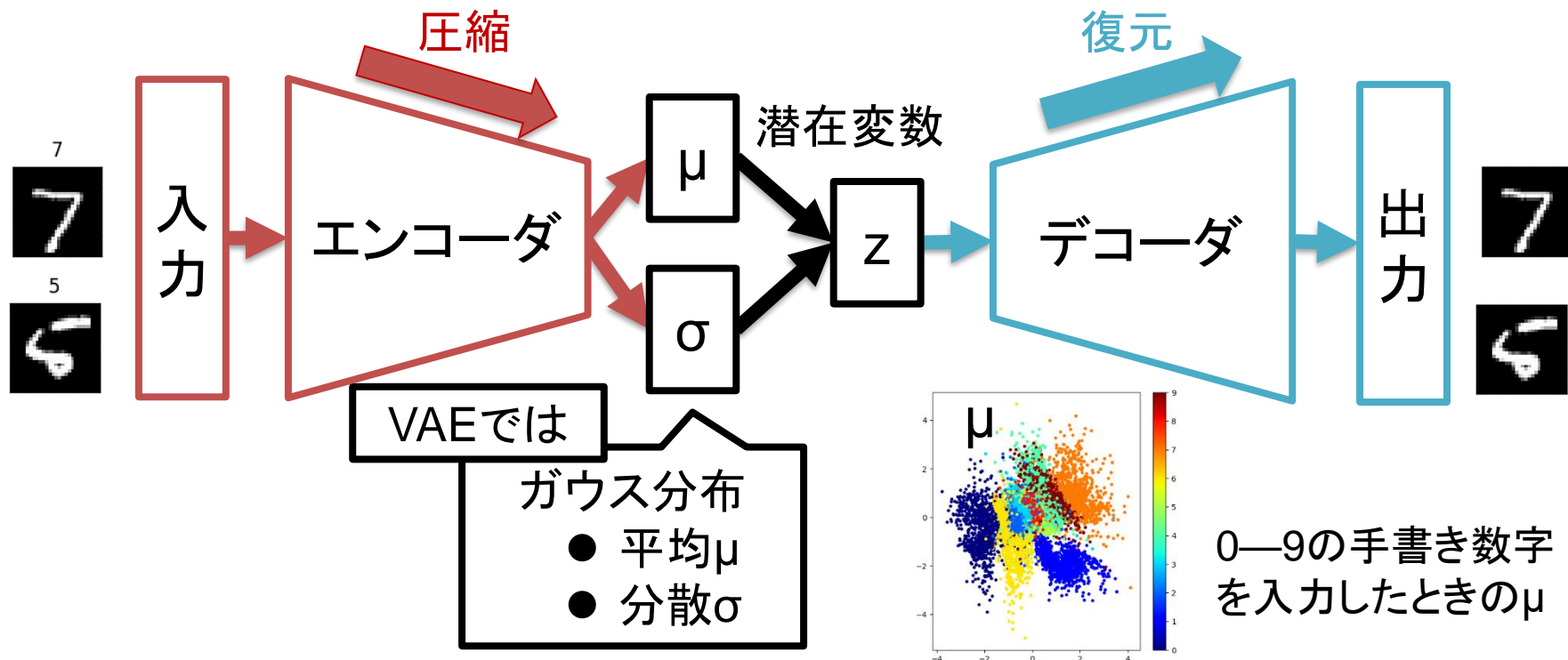
ラベル付けのコストが不要。

人が答えを知らない場合にも使える。



# Variational オートエンコーダ (VAE)

深層教師なし学習VAE(Kingma&Welling'14)は、入力をエンコーダで $\mu$ 、 $\sigma$ に圧縮し、ガウス分布に従う $z$ を生成する。 $z$ からデコーダで元のデータ形状に復元する。



損失関数は復元誤差と正則化項の和  
対数尤度をとって、損失関数を最適化させるように学習する。

$$\mathcal{L} = E - D_{\text{KL}}$$

- 復元誤差 入力と出力を近づける
- 正則化項  $\mu \rightarrow 0$ 、 $\sigma \rightarrow 1$ に近づける

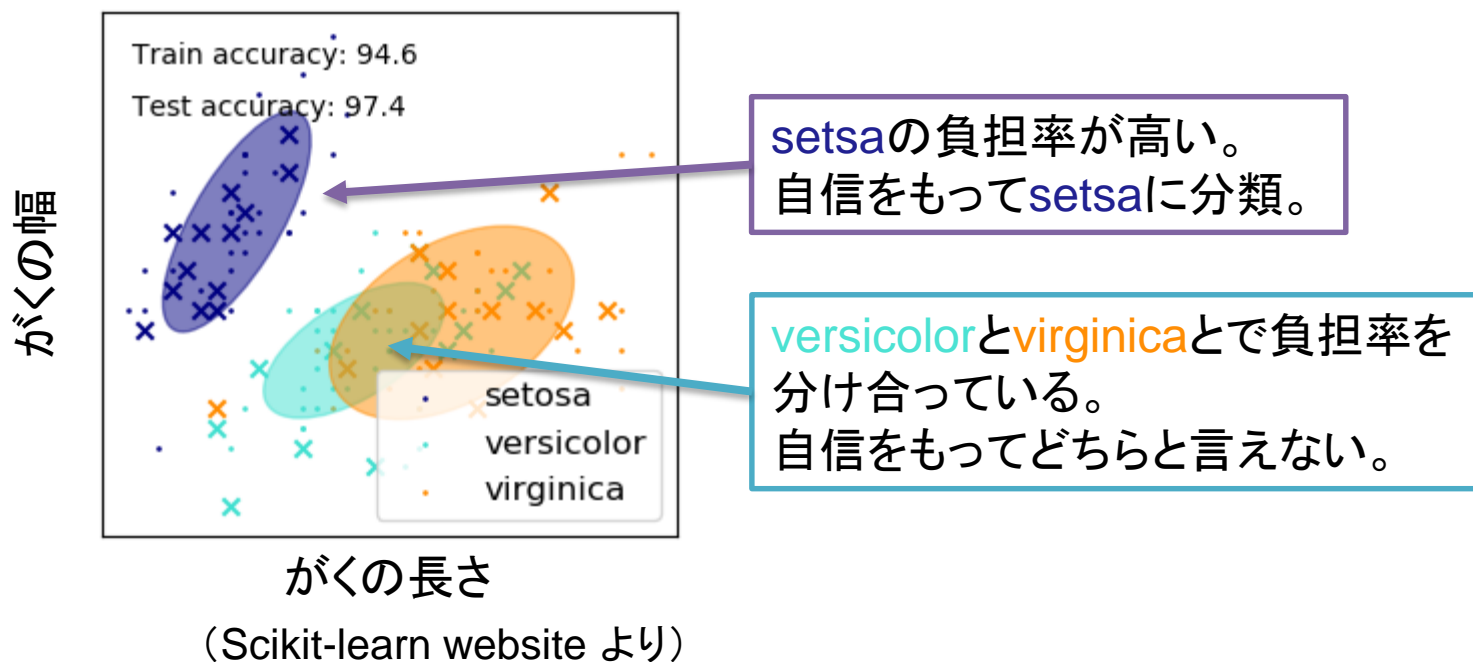
# 混合ガウスモデル (GMM)

クラスタリング データ点を複数のクラスに自動で分類する手法

- 混合ガウスモデルは、データ点の分布を複数のガウス分布で近似するクラスタリングの一種。

クラス間に境界線を引くのではなく、いくつかのガウス分布が重なっている範囲は、それぞれのガウス分布の負担率で表される(ソフトクラスタリング)。

アヤメ・データの例

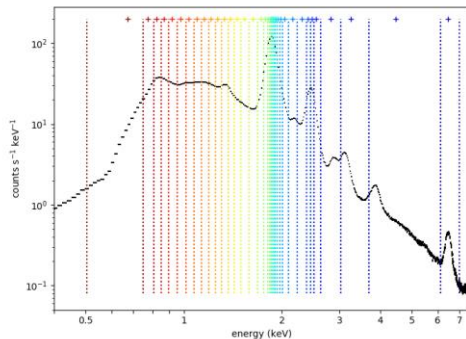


# Tycho's SNRへ機械学習を応用

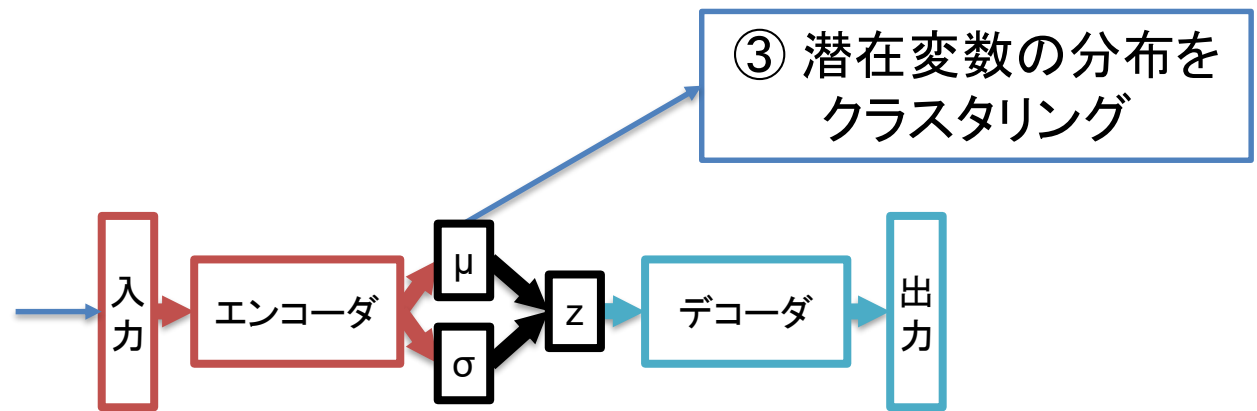
オートエンコーダ(要素抽出、次元削減)とクラスタリング(分類)を組み合わせると非線形で高次元の複雑なデータを自動で分類できると考案

→ Chandra衛星によるTycho's SNRの観測データに適用

- ① Tycho全体のスペクトルを、自動で37分割
- ② 画像の各ピクセルのスペクトルをVAEで次元削減(37次元→4次元)
- ③ 混合ガウスモデル(4次元ガウシアン8個)でクラスタリング



① 観測データ

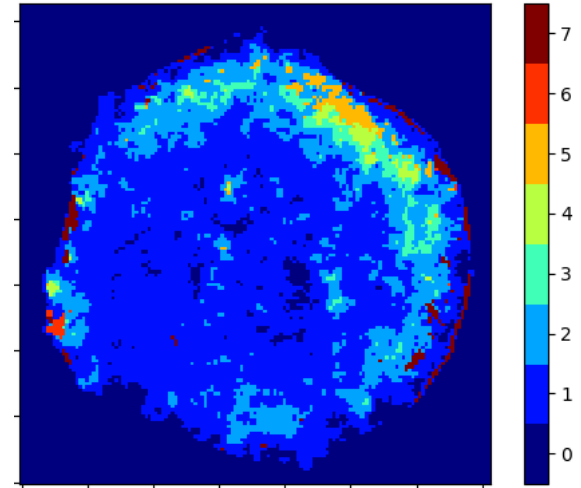
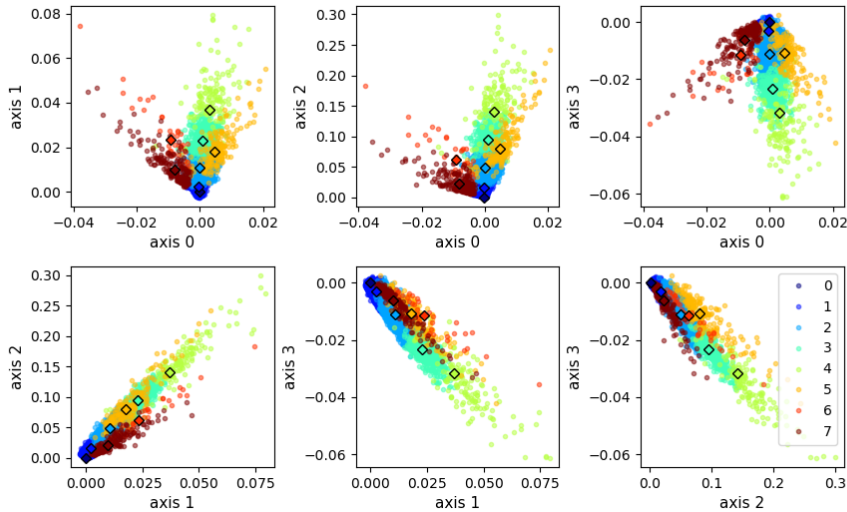


② VAEで次元削減

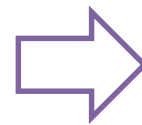


# Tycho を機械学習した結果

画像のピクセルごとのスペクトルを機械学習で次元削減・クラスタリングした結果



VAEの潜在変数(4次元)を2次元ずつ  
組合わせて図示  
クラスタリングによるクラスごとに色分け

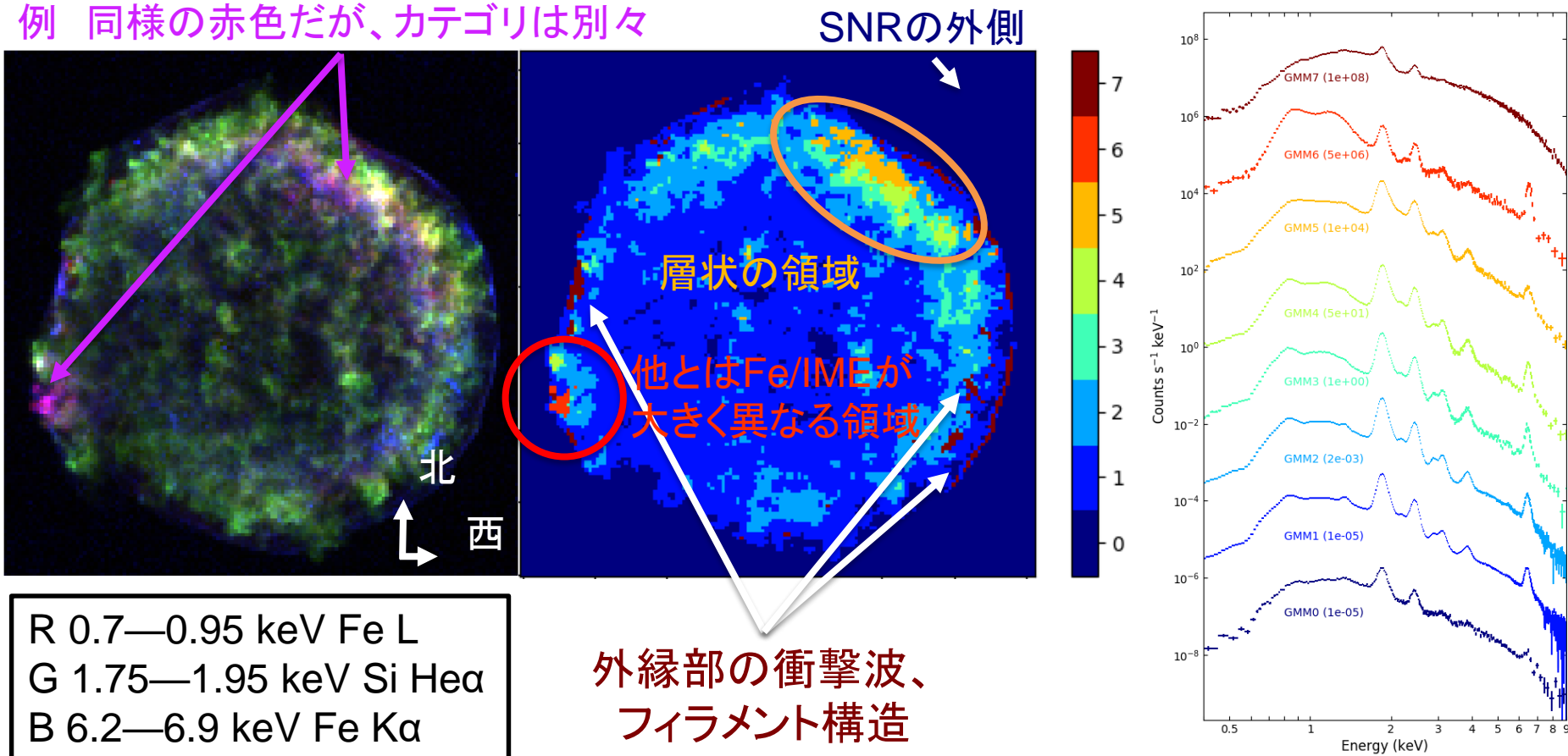


各データ点を画像上の位置に戻すと、  
クラス分けしたマップができる

# 3色カラーイメージ vs クラスタリングマップ

- ✓ 機械学習によって、特徴的な構造(既知のものも含めて)を分類できている
- ✓ 3色画像では同じような色でも、機械学習では別カテゴリに分類できている
- ✓ 機械学習では空間情報を与えていないが、空間的な分類が自動でされた

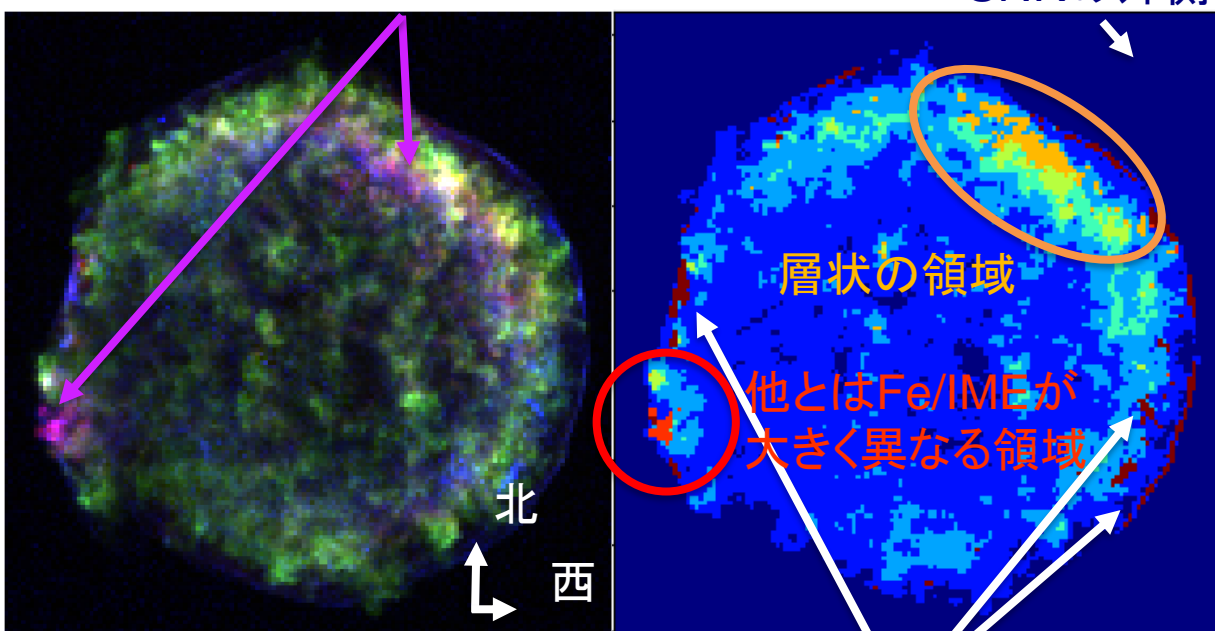
例 同様の赤色だが、カテゴリは別々



# 3色カラーイメージ vs クラスタリングマップ

- ✓ 機械学習によって、特徴的な構造(既知のものも含めて)を分類できている
- ✓ 3色画像では同じような色でも、機械学習では別カテゴリに分類できている
- ✓ 機械学習では空間情報を与えていないが、空間的な分類が自動でされた

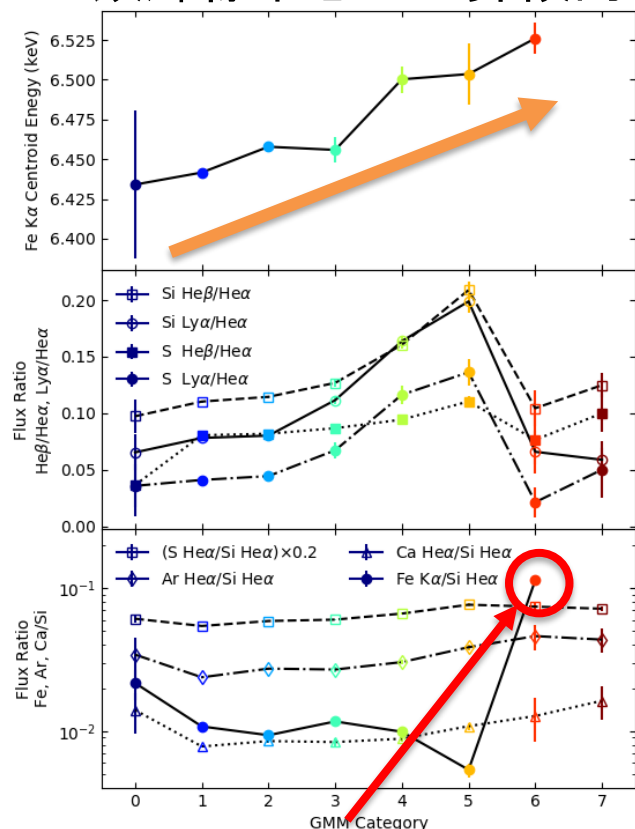
例 同様の赤色だが、カテゴリは別々



R 0.7—0.95 keV Fe L  
 G 1.75—1.95 keV Si Heα  
 B 6.2—6.9 keV Fe Kα

外縁部の衝撃波、  
 フィラメント構造

鉄輝線中心は上昇傾向



Fe/Si が突出

## 結果のまとめ

---

- 機械学習 (VAE、GMM) を Tycho's SNR の X線解析に応用。スペクトルの特徴的な領域が 自動的に分類 された。
  - 衝撃波、フィラメント構造 (ベキ型放射)
  - 南東の鉄ノット
  - 北西のイジェクタの電離状態の変化
  - 機械学習では空間情報を与えていないが、空間的な分類が自動的にされた
- オートエンコーダ (次元削減) とクラスタリング を組み合わせた手法は観測データから 特徴を抽出し分類する強力なツール となりうる。
  - 教師なし学習であるため、データへのラベル付けは不要。
  - 教師なし学習では何が学習されるかわからない。人が気づいていない特徴を抜き出す可能性もある。
  - 自動的に興味深い特徴を抽出することで詳細解析すべきところを教えてください。

# 今後の改善点

---

- 深層学習モデル VAE
  - VAEは潜在変数に0を中心とした多次元ガウス分布を仮定するため、特徴量空間でのデータ点は0を中心とした単峰な分布をとる。そのため、0付近のデータ点が集中するところではクラスタリングしづらい。
  - WAE、GMMVAEなどのVAEの改良版で改善できる可能性がある
  - 実質的に教師なしで分類をできるようなモデルも出てきているので、試す価値はありそう
- 解析済みのスペクトルやシミュレーションを学習データとして用いて、教師あり学習の手法で同様のスペクトル分類器を作ることもできる
- 1個1個のスペクトルだけでなく、空間情報(隣接するスペクトルの情報も同時に使うなど)も利用すれば、特徴的な空間構造をより捉えられる可能性もある。

# 今後の展開

---

- 教師なし機械学習による観測データからの特徴抽出・分類は、観測データに含まれる特徴を把握する強力なツールとなりうる。
- 本手法は1個1個のスペクトルを入力データとしているため、さまざまな観測に**応用可能**である。
  
- **様々な天体への応用**
  - 広がった天体の空間的な把握（銀河団、星形成領域など）
  - スペクトルの時間変化の把握
  
- **他波長の観測への応用**
  - 例えば、電波観測データは時間、空間、速度の次元を持っていて、データの次元がX線観測と似ている。