

4a-6-3

周期解析による変動天体検出

2023.3.6

千原直己(大阪大学), 高田唯史(国立天文台), 藤原靖宏(NTT), 鬼塚真(大阪大学)

目次

- 研究背景
- 研究方針
- 提案手法
 - 周期解析
 - 変動天体検出
- 実験

目次

- **研究背景**
- 研究方針
- 提案手法
 - 周期解析
 - 変動天体検出
- 実験

変動天体

変動の対象

- 光度
- 位置

変動の性質

- 一過的
- 周期的
- 局所的
- 大域的

変動天体

変動の対象

- 光度
- 位置

変動の性質

- 一過的
- 周期的
- 局所的
- 大域的

本研究の対象は

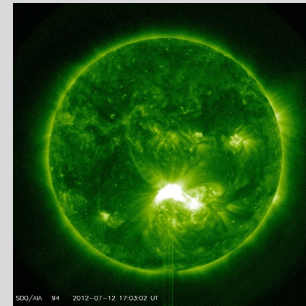
光度が周期的に変動する天体

変動天体

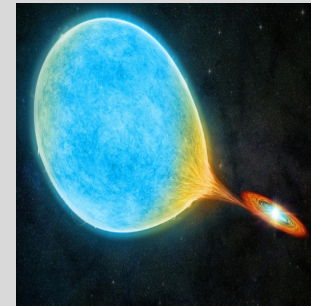
検出の利点

- 銀河系の形成や進化に関する研究への貢献
- ブラックホールの発見

画像例



閃光星^[1]



激変星^[2]

本研究の対象は

光度が周期的に変動する天体

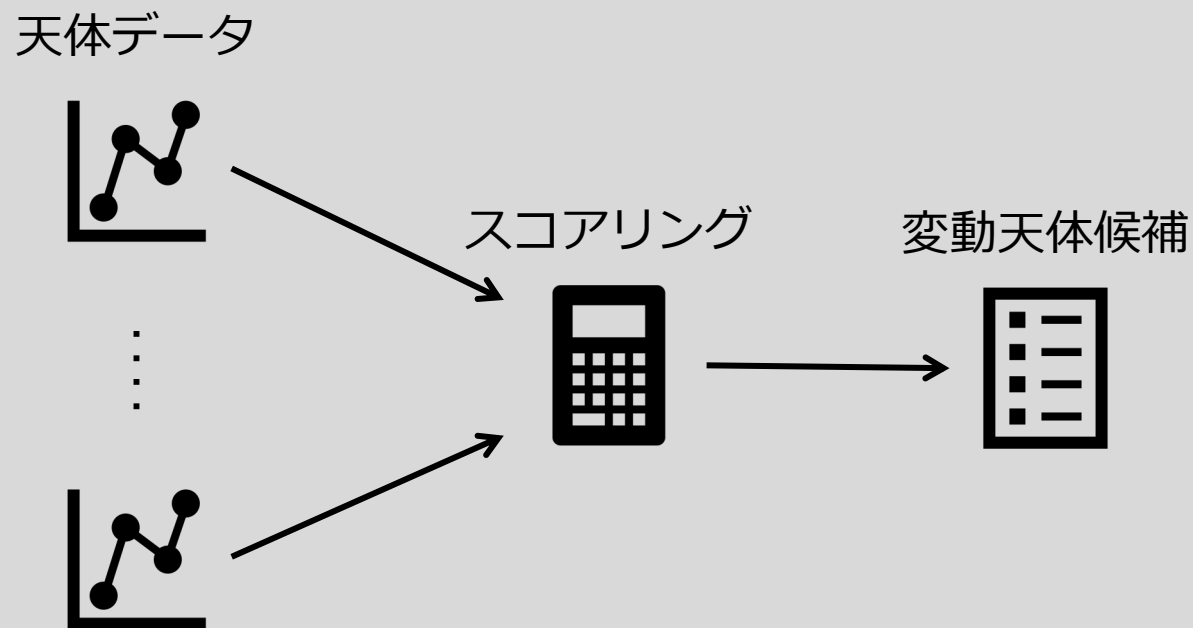
[1] <https://astronomy.com/news/videos/2014/03/first-sightings-of-solar-flare-phenomena-confirm-3-d-models-of-space-weather>

[2] <https://www.techexplorist.com/pair-stars-extremely-short-orbit/54131/>

既存技術

B.Sesarらの手法^[3]

統計的数値を活用したスコアリングを基にした変動天体検出手法

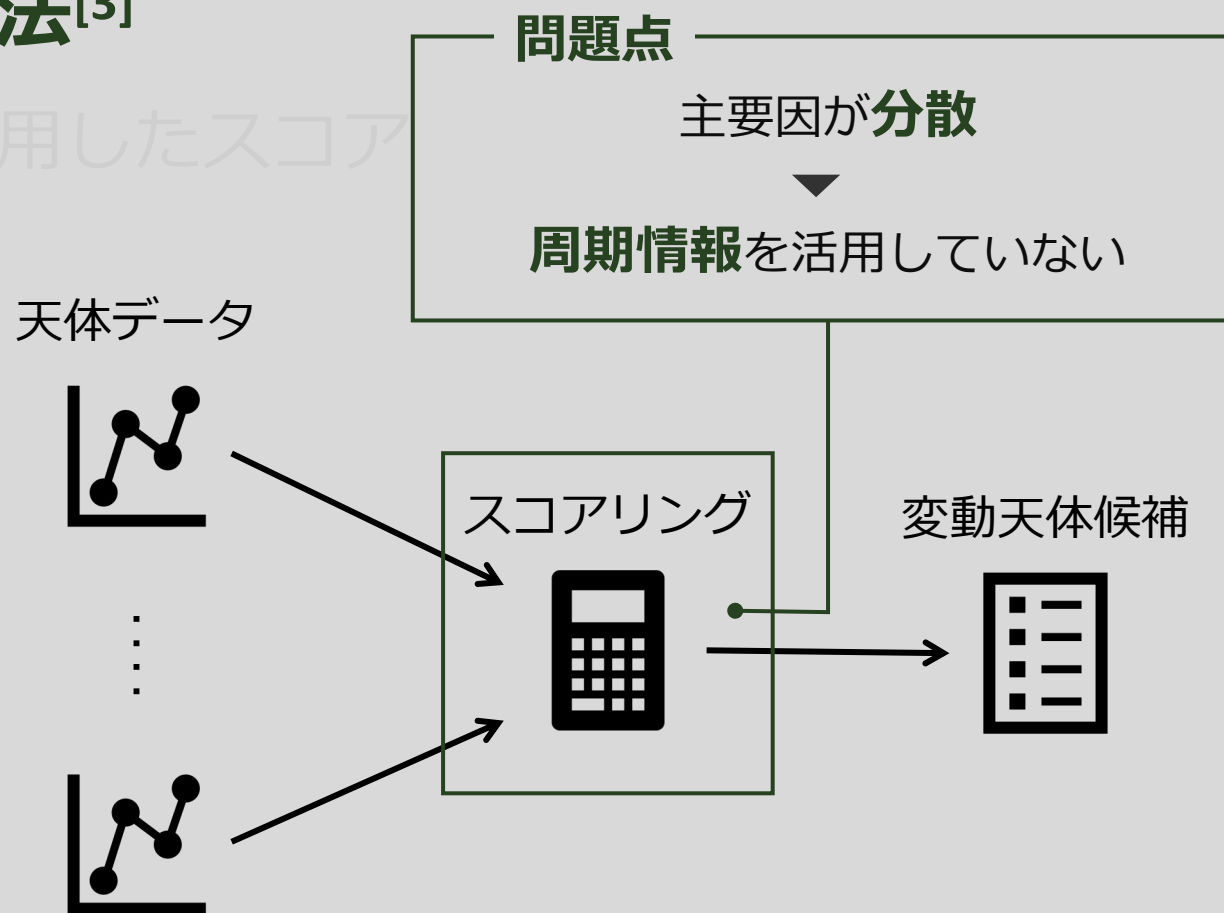


既存技術

B.Sesarらの手法^[3]

統計的数値を活用したスコア

検出手法



目次

- 研究背景
- **研究方針**
- 提案手法
 - 周期解析
 - 変動天体検出
- 実験

目的

目的

周期情報を活用した変動天体検出

目的

周期情報を活用した変動天体検出

- 天体データからの周期情報の抽出に**回帰分析**を活用する
- 変動天体候補セット取得のために, 抽出した周期情報を基に**分類タスク**を解決する

技術的課題

天体データのスパース性

- 天体データに多数の欠損値が混在している

特徴量候補の不足

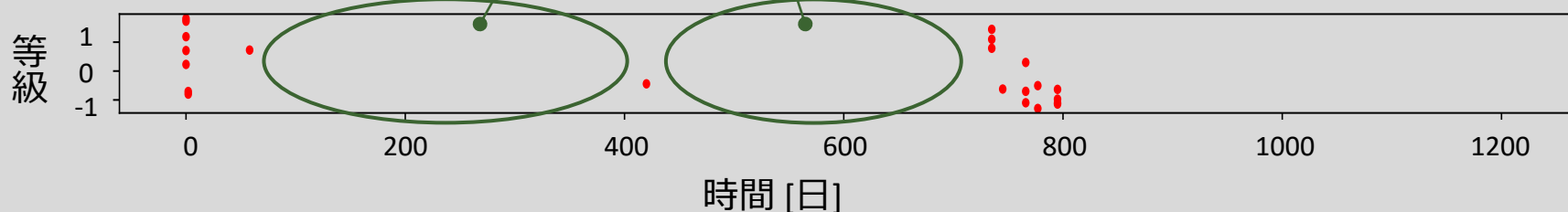
- 回帰分析によって得られる情報は周期情報のみ

技術的課題

天体データのスパース性

- 天体データに多数の欠損値が混在している

天体データ例



特徴量候補の不足

- 回帰分析によって得られる情報は周期情報のみ

技術的課題

天体データのスパース性

特徴量候補の不足

天体データに多数の欠損値

回帰分析に用いる保証が

欠損値が多いデータに対して
ナイーブな回帰分析を適用すると**過学習**が生じる

技術的課題

天体データのスパース性

- 天体データに多数の欠損値が混在している

特徴量候補の不足

- 回帰分析によって得られる情報は周期情報のみ



技術的課題

天体データのスパース性

- 天体データに多数の欠損値が混在している

特徴量候補の不足

- 回帰分析によって得られる情報は周期情報のみ



天体分野に特化した知見が予測に反映されない

技術的課題

天体データのスパース性

特徴量候補の不足

周期情報のみによる分類予測には
精度向上に**限界**がある

目次

- 研究背景
- 研究方針
- **提案手法**
 - 周期解析
 - 変動天体検出
- 実験

概要

スパースモデリングの活用

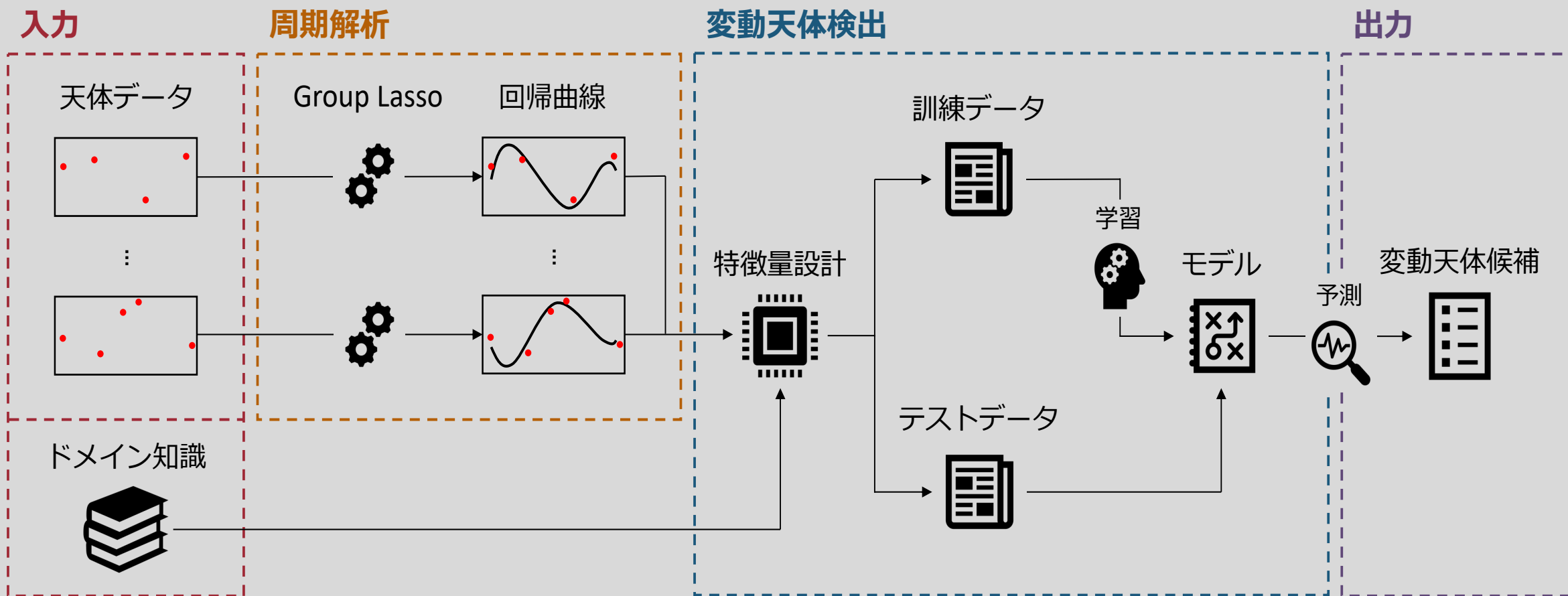
- 過学習の抑制を期待
- 本研究ではGroup Lassoを採択した

ドメイン知識の追加

- 天体分野の背景知識を反映させた予測モデルの実現

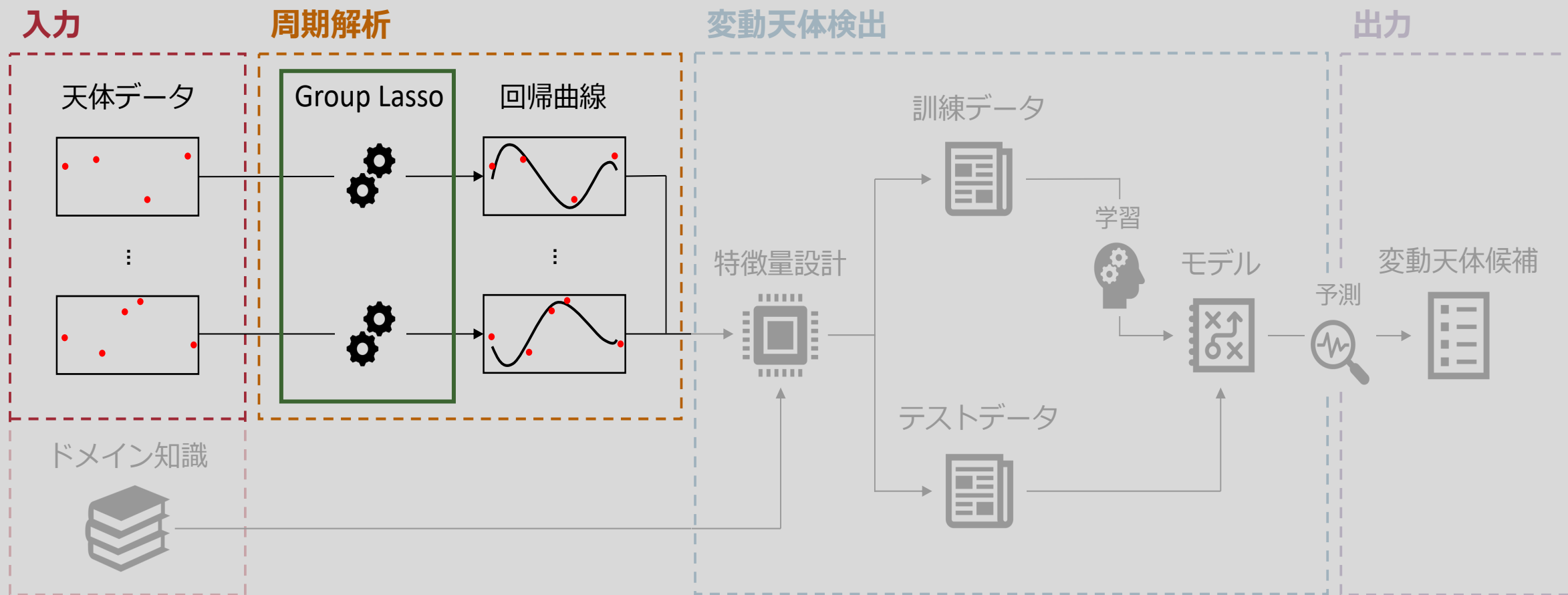
全体像

提案手法は2部構成



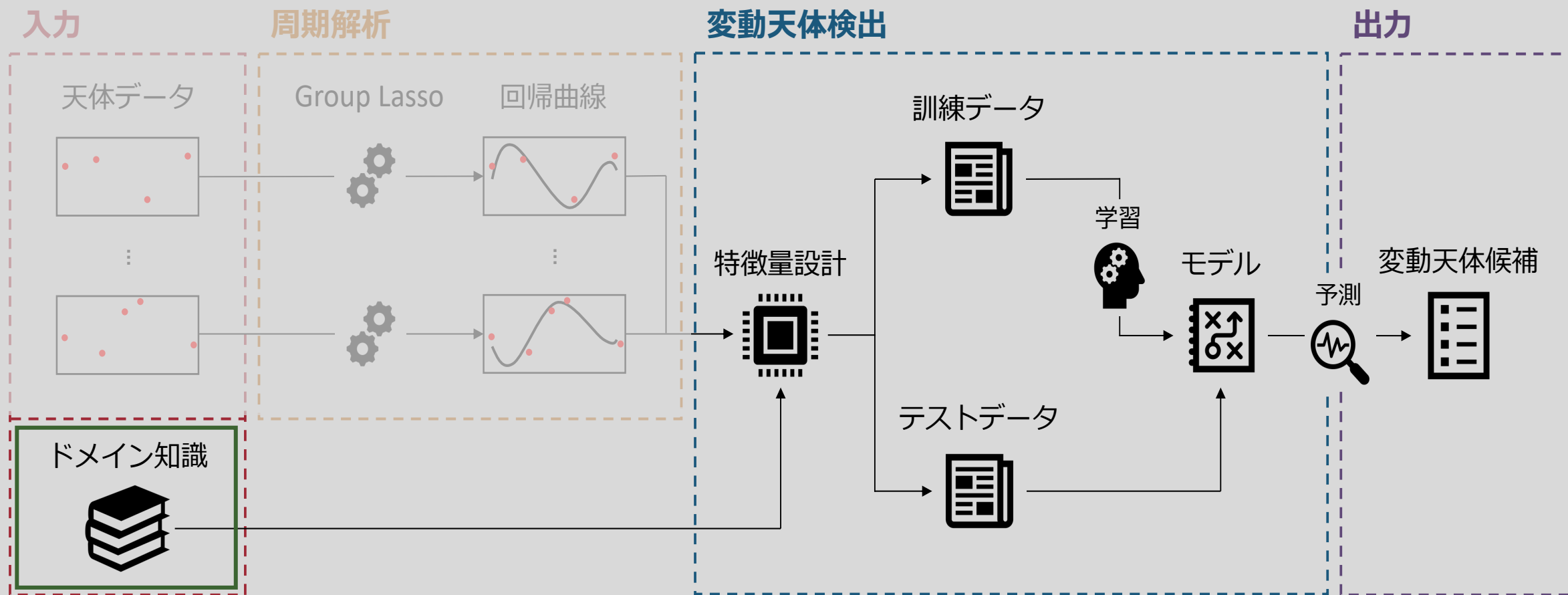
全体像

1. Group Lassoによる周期情報の抽出



全体像

2. 周期情報に加え, ドメイン知識を活用した予測



目次

- 研究背景
- 研究方針
- 提案手法
 - **周期解析**
 - 変動天体検出
- 実験

Group Lasso[4]

スパースモデリング

特徴量選択により入力の
重要部分のみを表現する技術



特徴

特徴量のグループ化により、
事前に把握している類似関係を
反映させた分析が可能



不要な特徴量グループ

(M_{ij} : i 番目グループに属する j 個目の特徴量)

Group Lasso[4]

スパースモデリング

特徴量選択により入力の
重要部分のみを表現する技術



過学習の抑制が期待できる

特徴

特徴量のグループ化により、
事前に把握している類似関係を
反映させた分析が可能



不要な特徴量グループ

(M_{ij} : i 番目グループに属する j 個目の特徴量)

Group Lasso[4]

スパースモデリング

特徴量選択により入力の
重要部分のみを表現する技術



過学習の抑制が期待できる

特徴

特徴量のグループ化により、
事前に把握している類似関係を
反映させた分析が可能



不要な特徴量グループ

(M_{ij} : i 番目グループに属する j 個目の特徴量)

回帰モデル

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sin(2\pi\nu_1 t_1) & \cdots & \sin(2\pi\nu_m t_1) & \cos(2\pi\nu_1 t_1) & \cdots & \cos(2\pi\nu_m t_1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sin(2\pi\nu_1 t_n) & \cdots & \sin(2\pi\nu_m t_n) & \cos(2\pi\nu_1 t_n) & \cdots & \cos(2\pi\nu_m t_n) \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_{2m} \end{pmatrix}$$

目的変数

説明変数

回帰係数

回帰モデル

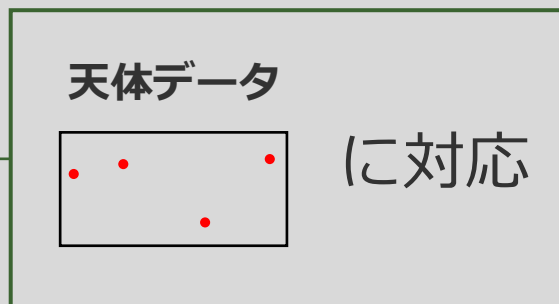
時刻 t_1 の観測データ点に対応

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sin(2\pi\nu_1 t_1) & \cdots & \sin(2\pi\nu_m t_1) & \cos(2\pi\nu_1 t_1) & \cdots & \cos(2\pi\nu_m t_1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sin(2\pi\nu_1 t_n) & \cdots & \sin(2\pi\nu_m t_n) & \cos(2\pi\nu_1 t_n) & \cdots & \cos(2\pi\nu_m t_n) \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_{2m} \end{pmatrix}$$

目的変数

説明変数

回帰係数



回帰モデル

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \overbrace{\sin(2\pi\nu_1 t_1) \cdots \sin(2\pi\nu_m t_1)}^{\text{正弦波}} & \overbrace{\cos(2\pi\nu_1 t_1) \cdots \cos(2\pi\nu_m t_1)}^{\text{余弦波}} \\ \vdots & \vdots \\ \sin(2\pi\nu_1 t_n) \cdots \sin(2\pi\nu_m t_n) & \cos(2\pi\nu_1 t_n) \cdots \cos(2\pi\nu_m t_n) \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_{2m} \end{pmatrix}$$

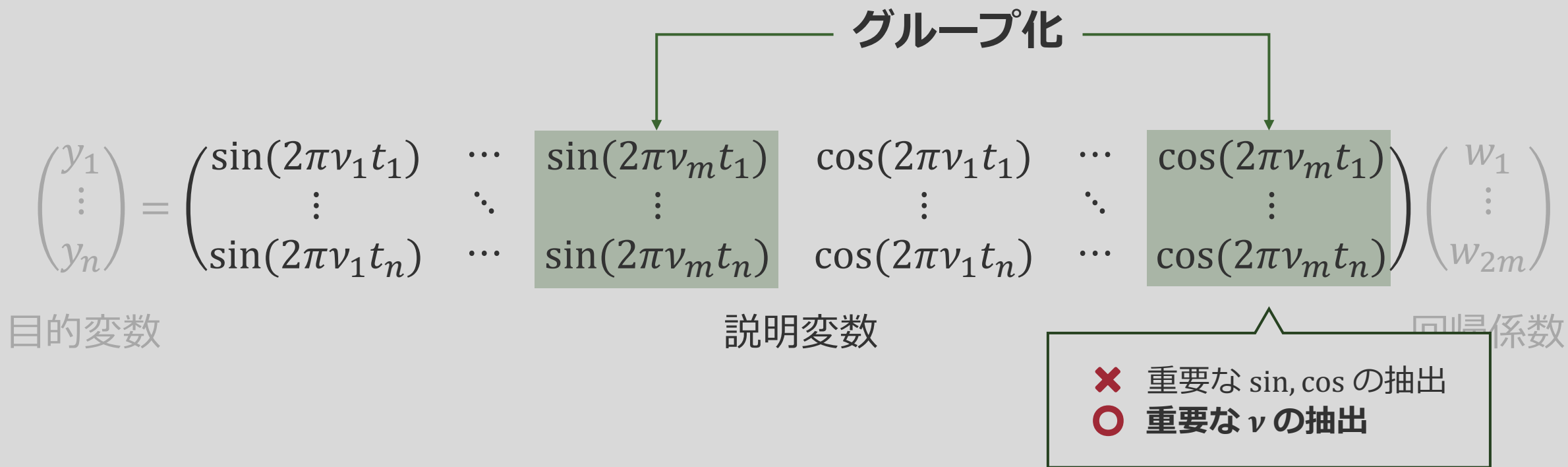
目的変数

説明変数

複数の周期長候補

回帰係数

回帰モデル



回帰モデル

$\sin(2\pi\nu_1 t)$ の**重要度**に対応

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sin(2\pi\nu_1 t_1) & \cdots & \sin(2\pi\nu_m t_1) & \cos(2\pi\nu_1 t_1) & \cdots & \cos(2\pi\nu_m t_1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sin(2\pi\nu_1 t_n) & \cdots & \sin(2\pi\nu_m t_n) & \cos(2\pi\nu_1 t_n) & \cdots & \cos(2\pi\nu_m t_n) \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_{2m} \end{pmatrix}$$

目的変数

説明変数

回帰係数

抽出したい**周期情報**

目次

- 研究背景
- 研究方針
- 提案手法
 - 周期解析
 - **変動天体検出**
- 実験

特徴量

周期情報

- `w` 周期解析により算出した回帰係数
- `reg_score` 回帰曲線による予測値と実際の目的変数の一致度

ドメイン知識

- `平均m_ap40` 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- `z_apertureflux_40_mag` ... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- `diff_ap40` `平均m_ap40` と `z_apertureflux_40_mag`の差分
- `data_size` 天体データに含まれる観測データ点の数

特徴量

周期情報

w 周期解析により算出した回帰係数

reg_score 回帰曲線による予測値と実際の目的変数の一致度

特徴量

周期情報

w 周期解析により算出した回帰係数

reg_score 回帰曲線による予測値と実際の目的変数の一致度

採択理由

reg_scoreは w の確らしさに相当する



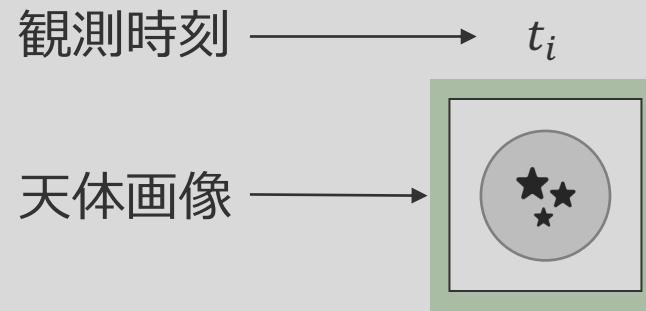
重みを考慮した予測が期待できる

特徴量

ドメイン知識

- 平均m_ap40 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- z_apertureflux_40_mag ... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- diff_ap40 平均m_ap40 と z_apertureflux_40_magの差分
- data_size 天体データに含まれる観測データ点の数

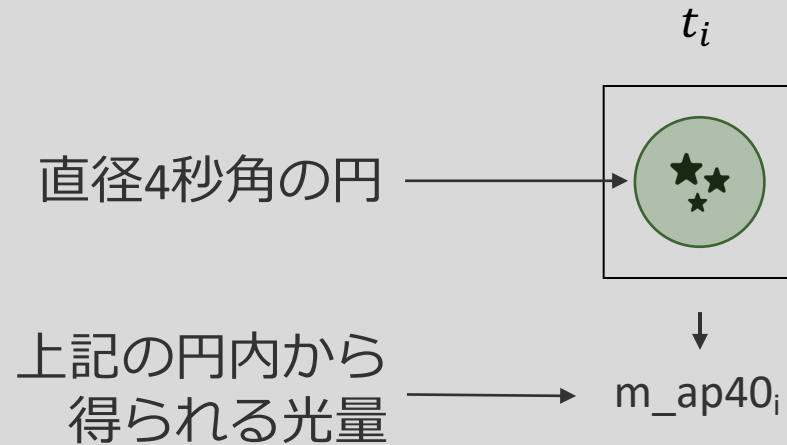
特徴量



ドメイン知識

- 平均m_ap40 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- z_apertureflux_40_mag ... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- diff_ap40 平均m_ap40 と z_apertureflux_40_magの差分
- data_size 天体データに含まれる観測データ点の数

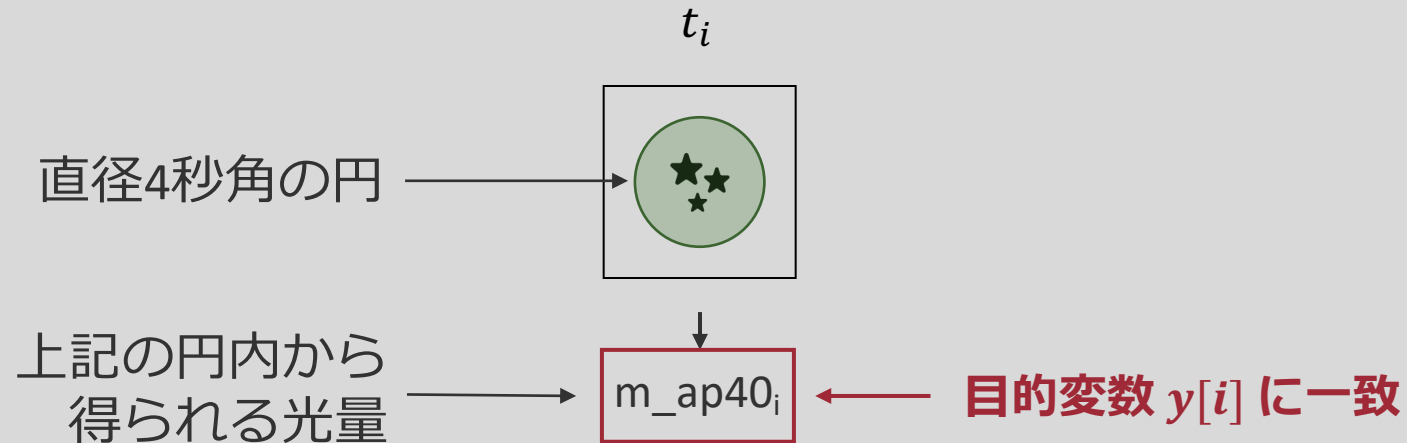
特徴量



ドメイン知識

- 平均 m_{ap40} 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- $z_{apertureflux_40_mag}$... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- $diff_ap40$ 平均 m_{ap40} と $z_{apertureflux_40_mag}$ の差分
- $data_size$ 天体データに含まれる観測データ点の数

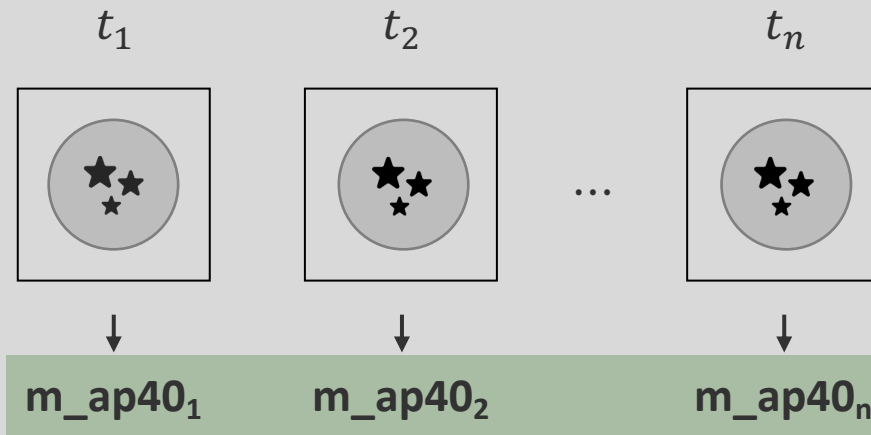
特徴量



ドメイン知識

- 平均 m_ap40 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- $z_apertureflux_40_mag$... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- $diff_ap40$ 平均 m_ap40 と $z_apertureflux_40_mag$ の差分
- $data_size$ 天体データに含まれる観測データ点の数

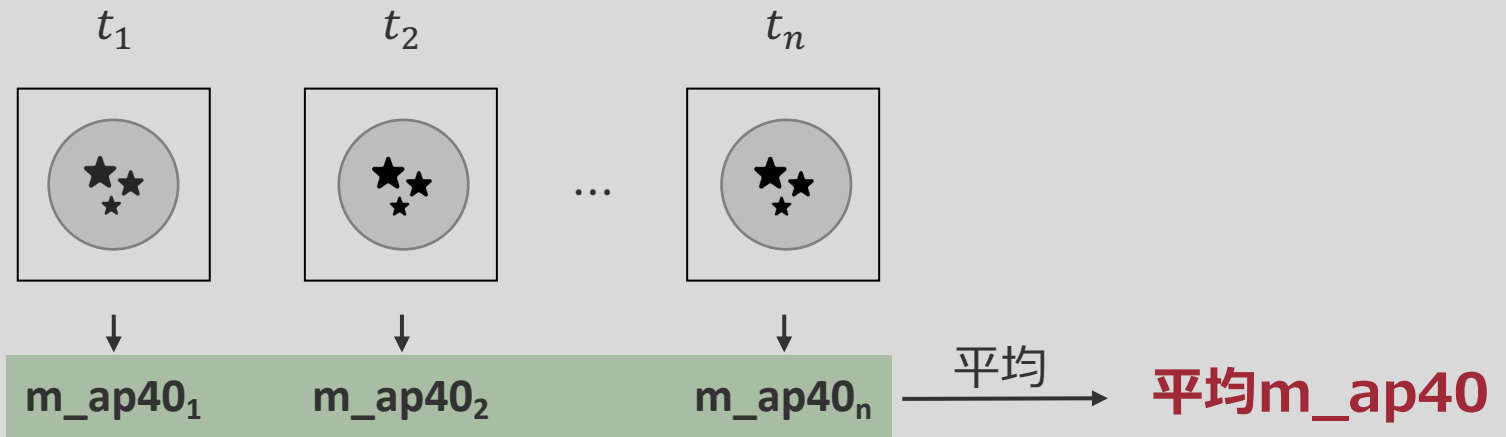
特徴量



ドメイン知識

- 平均 m_{ap40} 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- $z_{apertureflux_40_mag}$... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- $diff_ap40$ 平均 m_{ap40} と $z_{apertureflux_40_mag}$ の差分
- $data_size$ 天体データに含まれる観測データ点の数

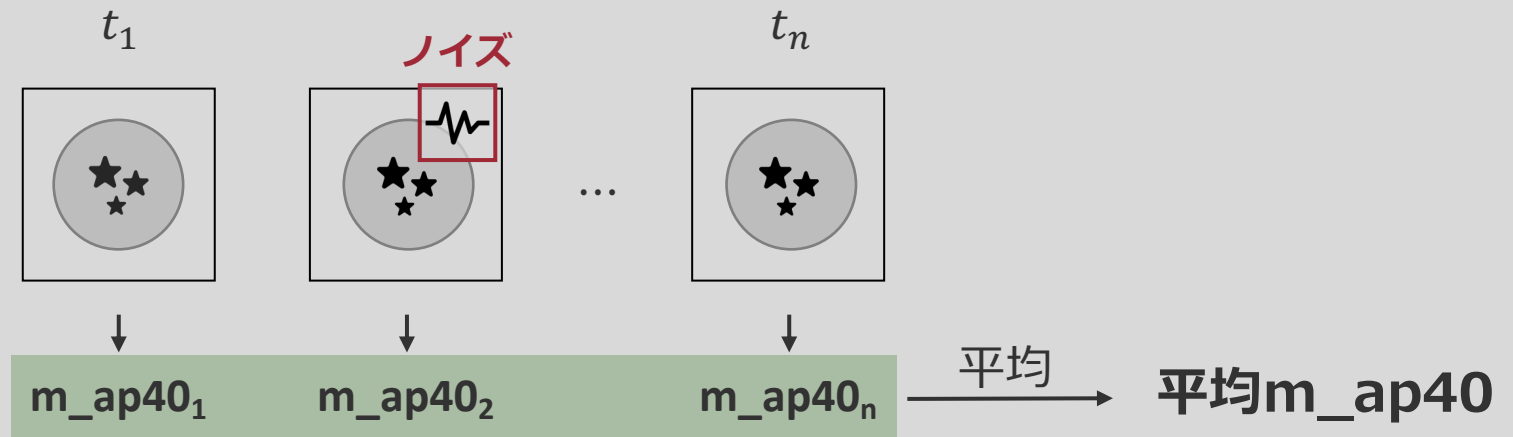
特徴量



ドメイン知識

- 平均 m_{ap40} 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- $z_{apertureflux_40_mag}$... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- $diff_ap40$ 平均 m_{ap40} と $z_{apertureflux_40_mag}$ の差分
- $data_size$ 天体データに含まれる観測データ点の数

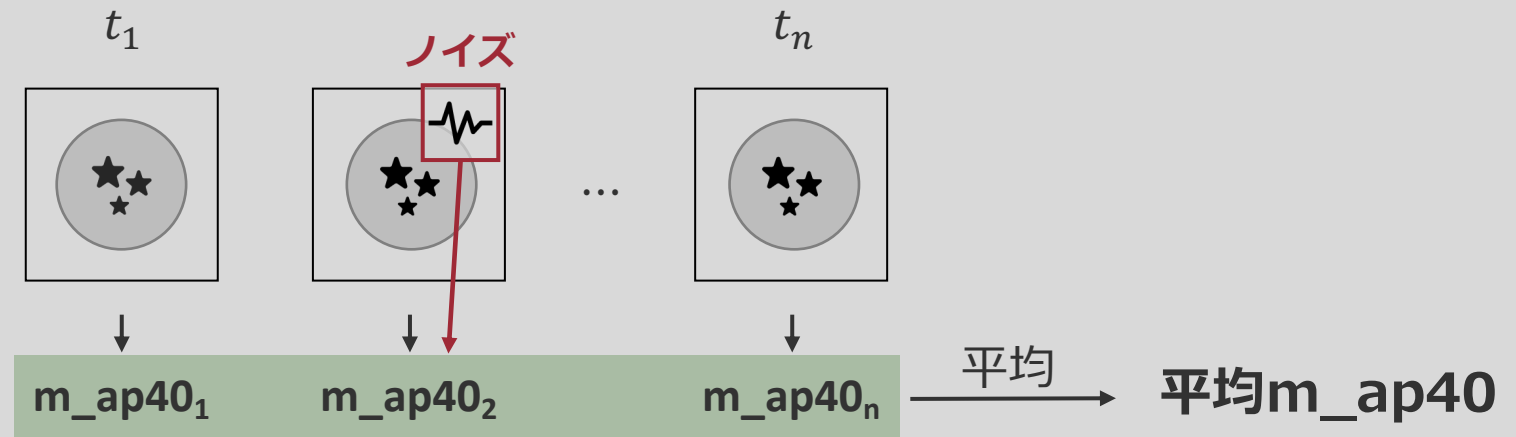
特徴量



ドメイン知識

- 平均 m_{ap40} 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- $z_{apertureflux_40_mag}$... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- $diff_ap40$ 平均 m_{ap40} と $z_{apertureflux_40_mag}$ の差分
- $data_size$ 天体データに含まれる観測データ点の数

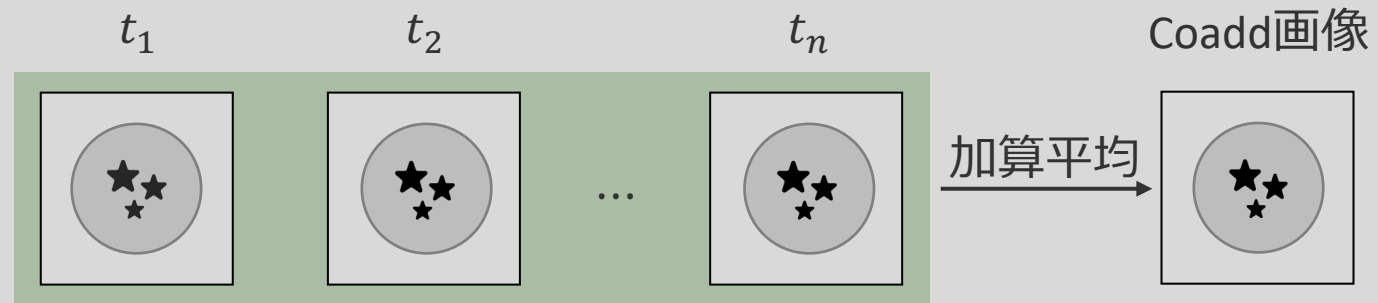
特徴量



ドメイン知識

- 平均 m_{ap40} 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- $z_{apertureflux_40_mag}$... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- $diff_{ap40}$ 平均 m_{ap40} と $z_{apertureflux_40_mag}$ の差分
- $data_size$ 天体データに含まれる観測データ点の数

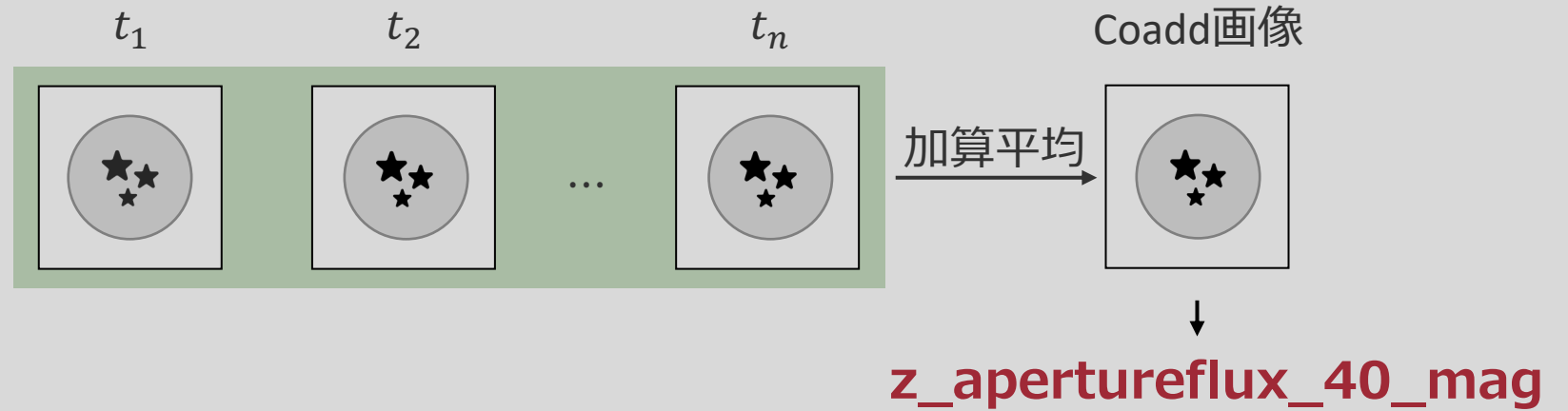
特徴量



ドメイン知識

- 平均m_ap40 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- z_apertureflux_40_mag ... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- diff_ap40 平均m_ap40 と z_apertureflux_40_magの差分
- data_size 天体データに含まれる観測データ点の数

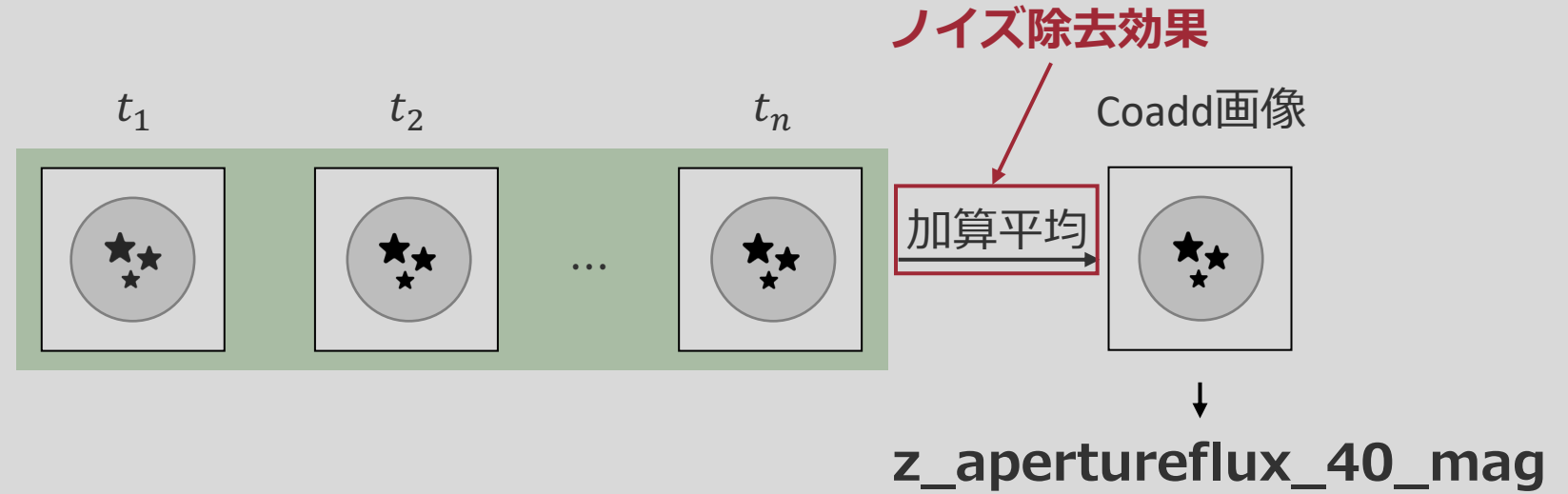
特徴量



ドメイン知識

- 平均m_ap40 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- z_apertureflux_40_mag ... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- diff_ap40 平均m_ap40 と z_apertureflux_40_magの差分
- data_size 天体データに含まれる観測データ点の数

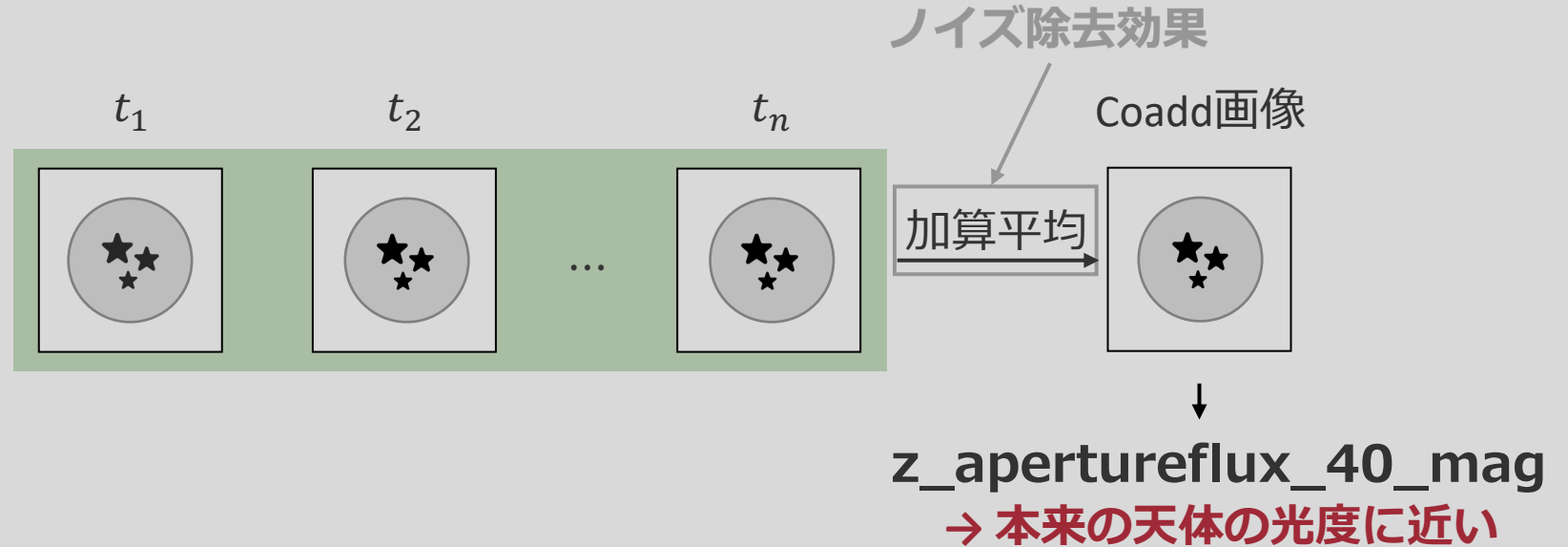
特徴量



ドメイン知識

- 平均m_ap40 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- z_apertureflux_40_mag ... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- diff_ap40 平均m_ap40 と z_apertureflux_40_magの差分
- data_size 天体データに含まれる観測データ点の数

特徴量



ドメイン知識

- 平均m_ap40 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- z_apertureflux_40_mag** ... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- diff_ap40 平均m_ap40 と z_apertureflux_40_magの差分
- data_size 天体データに含まれる観測データ点の数

特徴量

$$\text{diff_ap40} = \text{平均m_ap40} - \text{z_apertureflux_40_mag}$$

ドメイン知識

- 平均m_ap40 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- z_apertureflux_40_mag ... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- diff_ap40 平均m_ap40 と z_apertureflux_40_magの差分
- data_size 天体データに含まれる観測データ点の数

特徴量

$$\text{diff_ap40} = \text{平均m_ap40} - \text{z_apertureflux_40_mag}$$

目的変数の平均値天体本来の光度

ドメイン知識

- 平均m_ap40 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- z_apertureflux_40_mag ... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- diff_ap40 平均m_ap40 と z_apertureflux_40_magの差分
- data_size 天体データに含まれる観測データ点の数

特徴量

$$\text{diff_ap40} = \text{平均m_ap40} - \text{z_apertureflux_40_mag}$$

→ 天体データが受けている
ノイズの影響度

目的変数の平均値

天体本来の光度

ドメイン知識

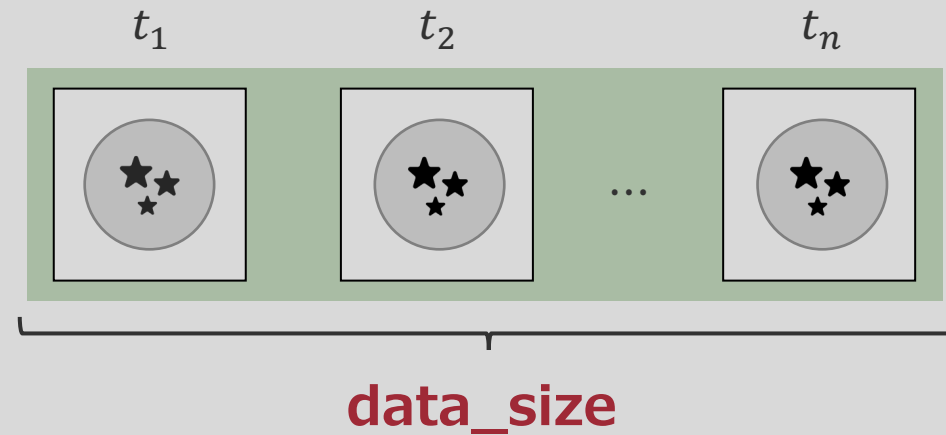
平均m_ap40 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値

z_apertureflux_40_mag ... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量

diff_ap40 平均m_ap40 と z_apertureflux_40_magの差分

data_size 天体データに含まれる観測データ点の数

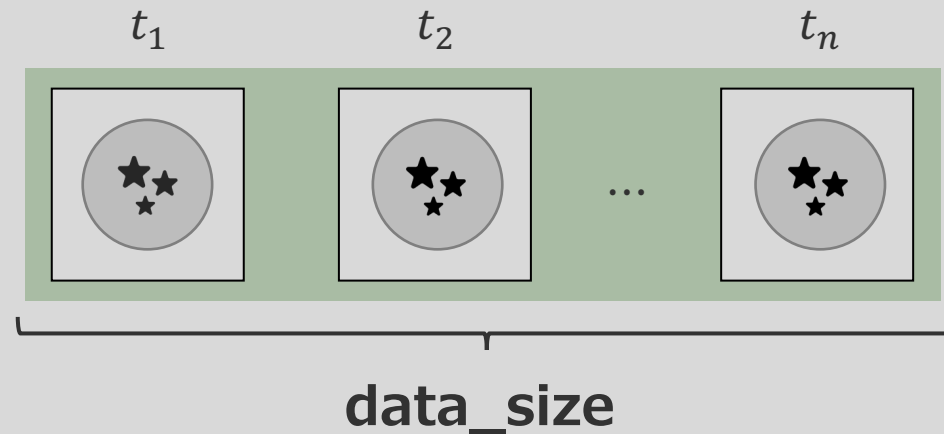
特徴量



ドメイン知識

- 平均m_ap40 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- z_apertureflux_40_mag ... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- diff_ap40 平均m_ap40 と z_apertureflux_40_magの差分
- data_size 天体データに含まれる観測データ点の数

特徴量



→ 観測点が少ない程, 過学習が生じやすい

ドメイン知識

- 平均m_ap40 画像上の対象天体を中心とした直径4秒角の円内の光量の平均値
- z_apertureflux_40_mag ... Coadd画像上の対象天体を中心とした直径4秒角の円内の光量
- diff_ap40 平均m_ap40 と z_apertureflux_40_magの差分
- data_size 天体データに含まれる観測データ点の数

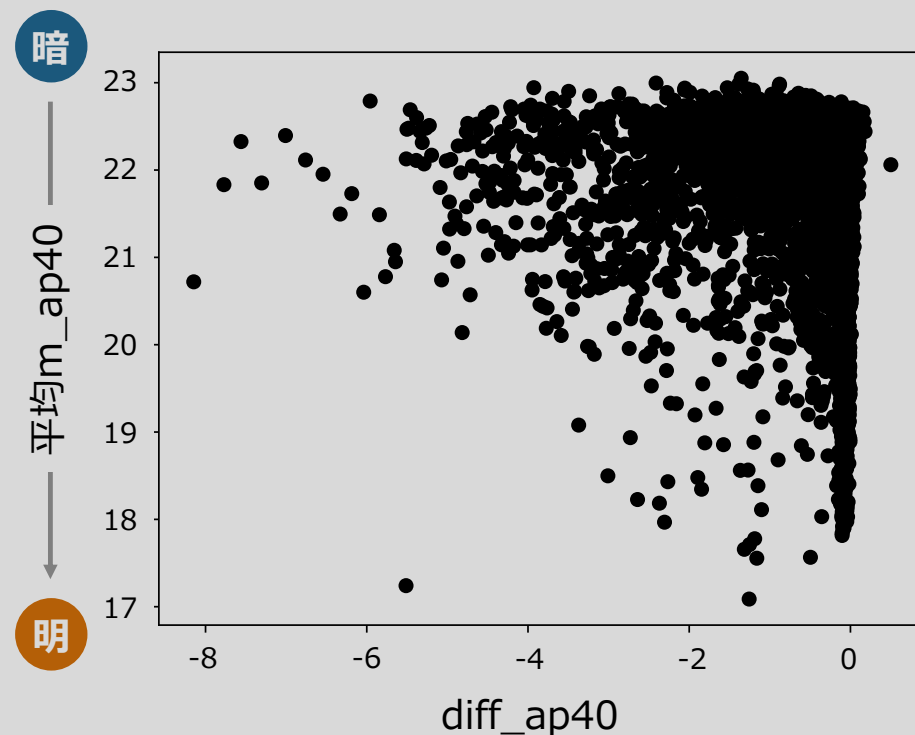
天体の光度とノイズ

背景

- **diff_ap40**はノイズの影響度に相当した特徴量
- 暗い天体ほどdiff_ap40の絶対値が大きい(右図参照)



diff_ap40 と 平均m_ap40 の関係図



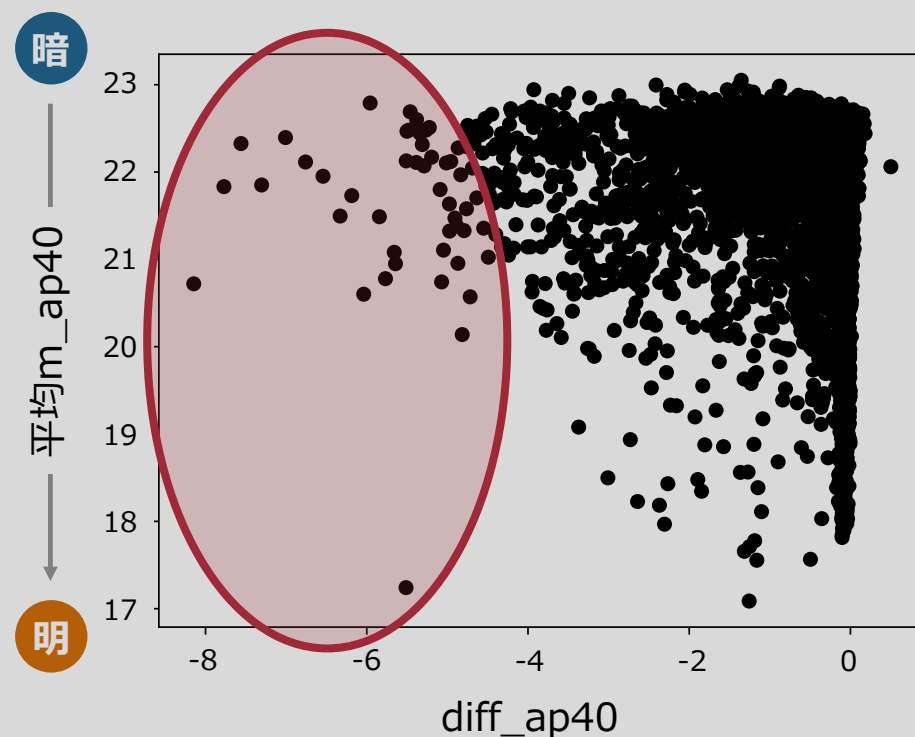
天体の光度とノイズ

背景

- **diff_ap40**はノイズの影響度に相当した特徴量
- 暗い天体ほどdiff_ap40の絶対値が大きい(右図参照)



diff_ap40 と 平均m_ap40 の関係図



天体の光度とノイズ

背景

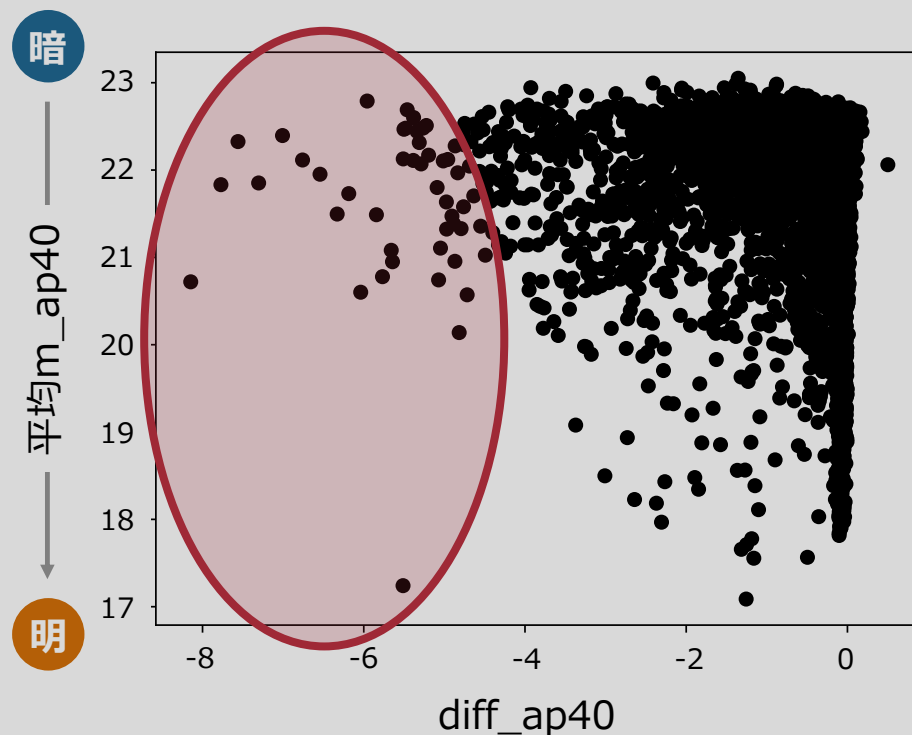
- **diff_ap40**はノイズの影響度に相当した特徴量
- 暗い天体ほどdiff_ap40の絶対値が大きい(右図参照)



暗い天体である程

ノイズの影響を強く 受ける

diff_ap40 と 平均m_ap40 の関係図



目次

- 研究背景
- 研究方針
- 提案手法
 - 周期解析
 - 変動天体検出
- **実験**

実験設定

目的

ベースライン

データセット

(実験1)

提案モデルの性能評価

B.Sesarらの手法^[3]

PDR2

(Public Data Release2)

(実験2)

周期解析の堅牢性評価

線形回帰

実データに基づいた
人工データ

実験設定

目的

ベースライン

データセット

(実験1)

提案モデルの性能評価

B.Sesarらの手法^[3]

PDR2
(Public Data Release2)

(実験2)

周期解析の堅牢性評価

線形回帰

実データに基づいた
人工データ

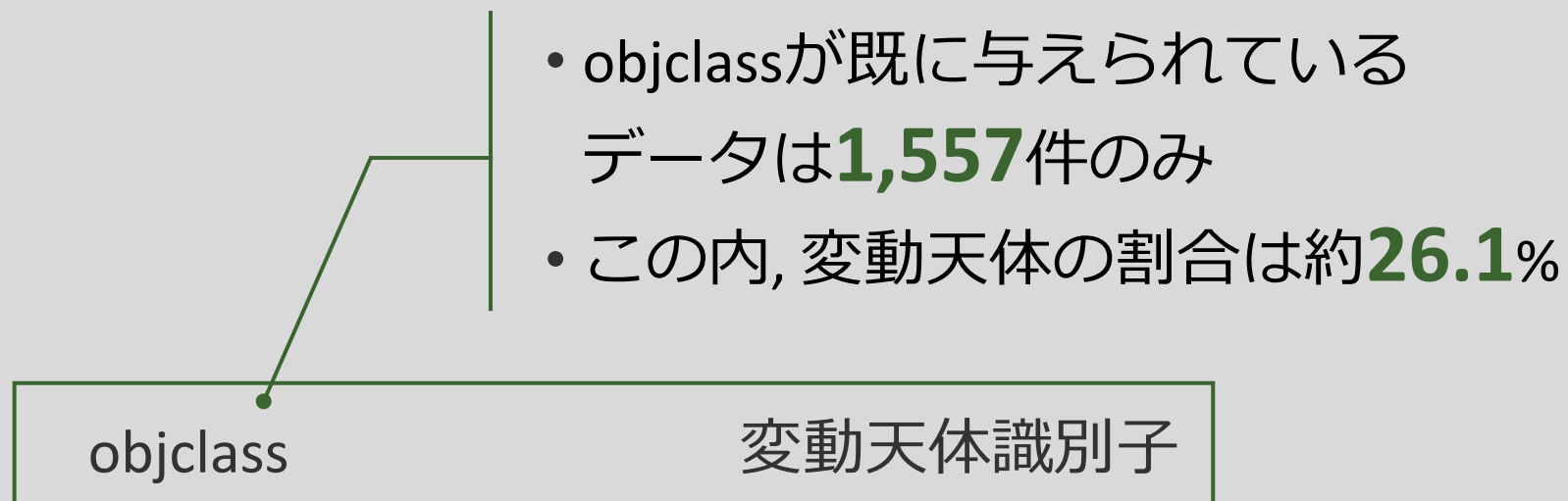
PDR2

以下のカラム構成を持つデータセット

名前	概要
objectid	天体識別子
mjd	観測時刻
m_ap40	天体光度
objclass	変動天体識別子

PDR2

以下のカラム構成を持つデータセット



提案モデルの性能評価

モデル	AUC	accuracy	precision	recall	F1
B.Sesar 手法	0.750	0.743	0.507	0.507	0.507
提案手法					
線形回帰 + ドメイン知識 無	0.633	0.729	0.432	0.109	0.168
線形回帰 + ドメイン知識 有	0.846	0.820	0.702	0.538	0.605
Lasso + ドメイン知識 無	0.854	0.825	0.769	0.478	0.586
Lasso + ドメイン知識 有	0.937	0.897	0.815	0.780	0.795
Group Lasso + ドメイン知識 無	0.860	0.828	0.793	0.476	0.590
Ours Group Lasso + ドメイン知識 有	0.939	0.899	0.820	0.785	0.801

提案モデルの性能評価

どの評価指標を確認しても, 大幅な性能向上が確認できる

モデル	AUC	accuracy	precision	recall	F1
B.Sesar 手法	0.750	0.743	0.507	0.507	0.507
提案手法					
線形回帰 + ドメイン知識 無	0.633	0.729	0.432	0.109	0.168
線形回帰 + ドメイン知識 有	0.846	0.820	0.702	0.538	0.605
Lasso + ドメイン知識 無	0.854	0.825	0.769	0.478	0.586
Lasso + ドメイン知識 有	0.937	0.897	0.815	0.780	0.795
Group Lasso + ドメイン知識 無	0.860	0.828	0.793	0.476	0.590
Ours Group Lasso + ドメイン知識 有	0.939	0.899	0.820	0.785	0.801

提案モデルの性能評価

スパースなデータから重要周期を抽出できている

モデル	AUC	accuracy	precision	recall	F1
B.Sesar 手法	0.750	0.743	0.507	0.507	0.507
提案手法					
線形回帰 + ドメイン知識 無	0.633	0.729	0.432	0.109	0.168
線形回帰 + ドメイン知識 有	0.846	0.820	0.702	0.538	0.605
Lasso + ドメイン知識 無	0.854	0.825	0.769	0.478	0.586
Lasso + ドメイン知識 有	0.937	0.897	0.815	0.780	0.795
Group Lasso + ドメイン知識 無	0.860	0.828	0.793	0.476	0.590
Ours Group Lasso + ドメイン知識 有	0.939	0.899	0.820	0.785	0.801

提案モデルの性能評価

天体の背景情報を正しく反映させた分類ができている

モデル	AUC	accuracy	precision	recall	F1
B.Sesar 手法	0.750	0.743	0.507	0.507	0.507
提案手法					
線形回帰 + ドメイン知識 無	0.633	0.729	0.432	0.109	0.168
線形回帰 + ドメイン知識 有	0.846	0.820	0.702	0.538	0.605
Lasso + ドメイン知識 無	0.854	0.825	0.769	0.478	0.586
Lasso + ドメイン知識 有	0.937	0.897	0.815	0.780	0.795
Group Lasso + ドメイン知識 無	0.860	0.828	0.793	0.476	0.590
Ours Group Lasso + ドメイン知識 有	0.939	0.899	0.820	0.785	0.801

周期解析の堅牢性評価

定性的な評価

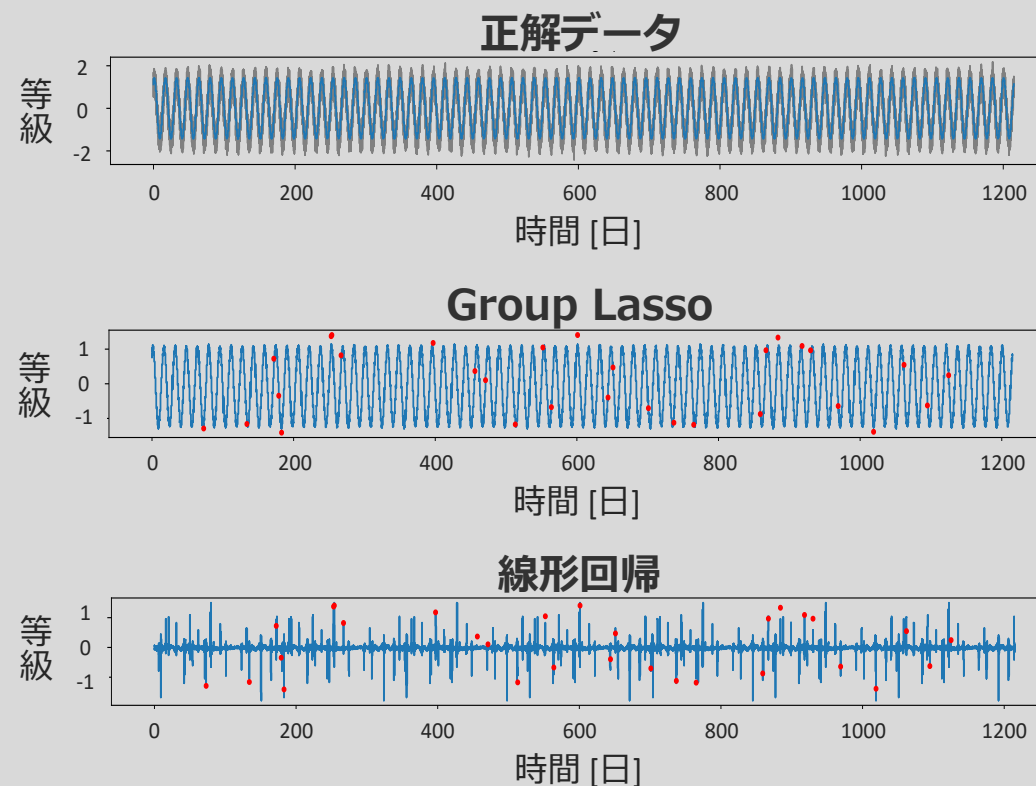
Group Lasso

- 正解データと類似
- 重要周期が明確

線形回帰

- 重要周期が不明
- 過学習の発生

周期解析結果



周期解析の堅牢性評価

定性的な評価

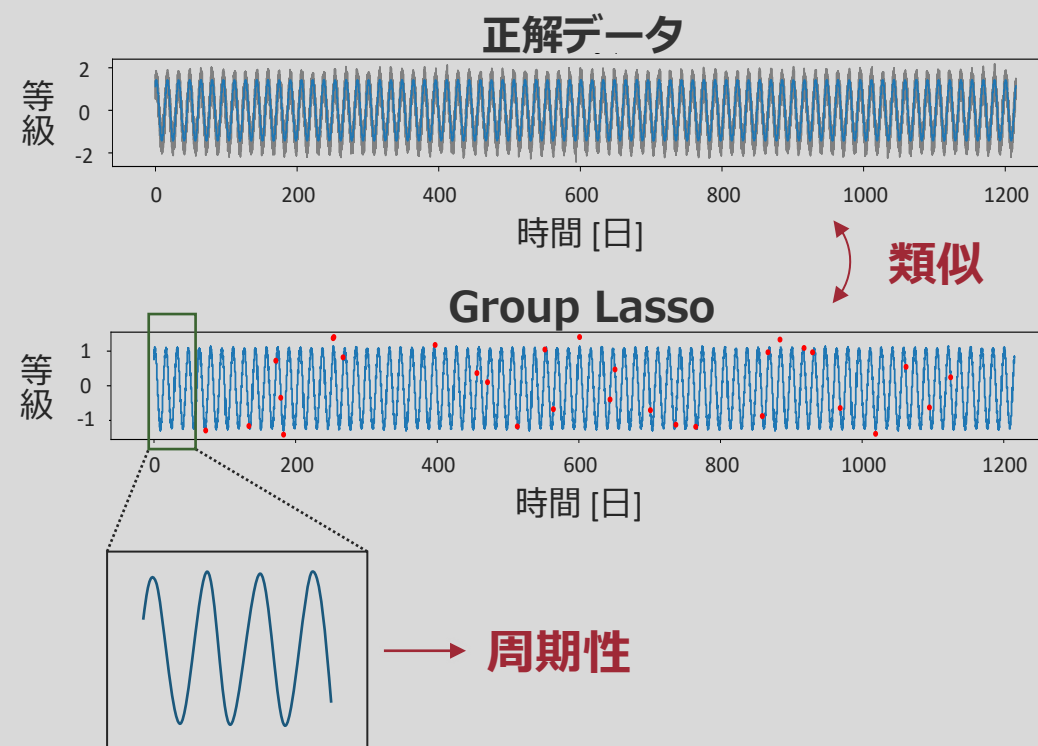
Group Lasso

- 正解データと類似
- 重要周期が明確

線形回帰

- 重要周期が不明
- 過学習が発生

周期解析結果



周期解析の堅牢性評価

定性的な評価

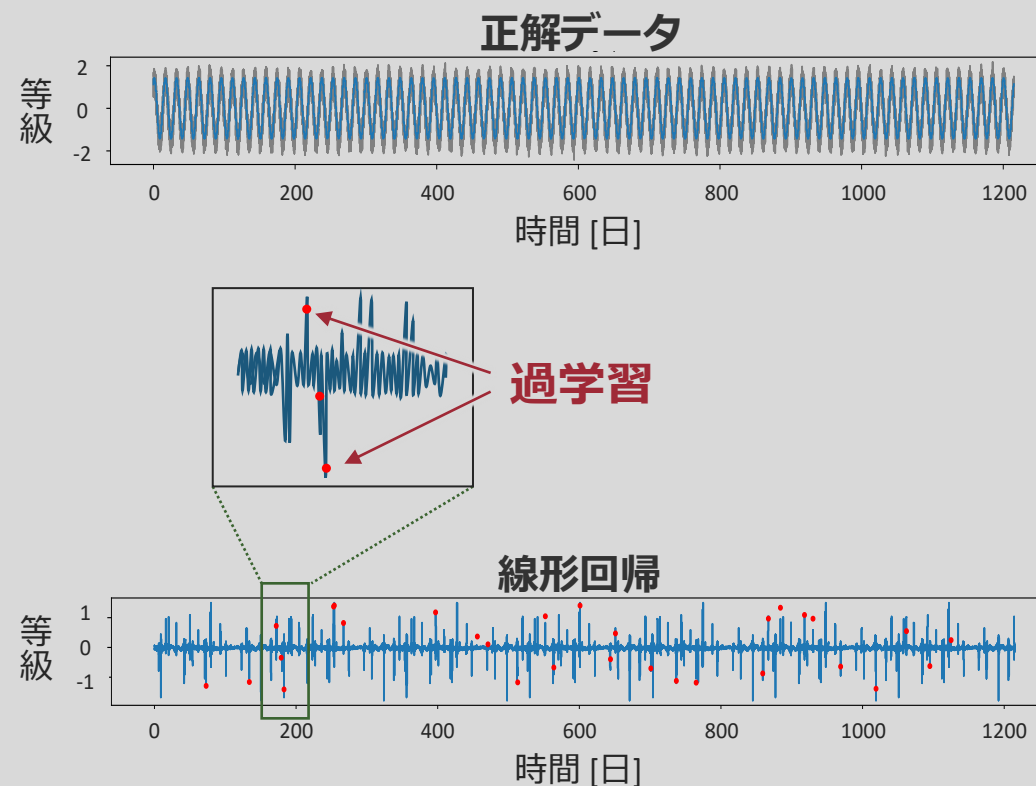
Group Lasso

- 正解データと類似
- 重要周期が明確

線形回帰

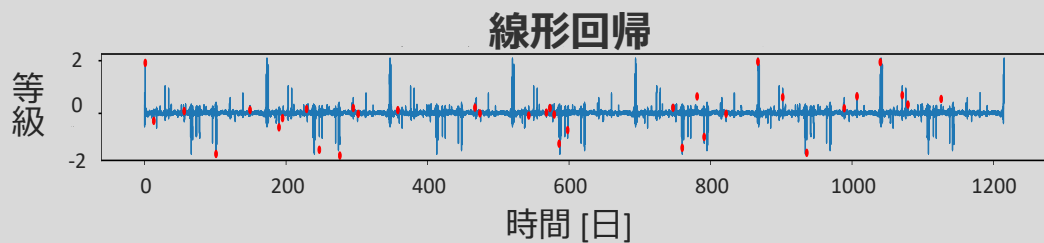
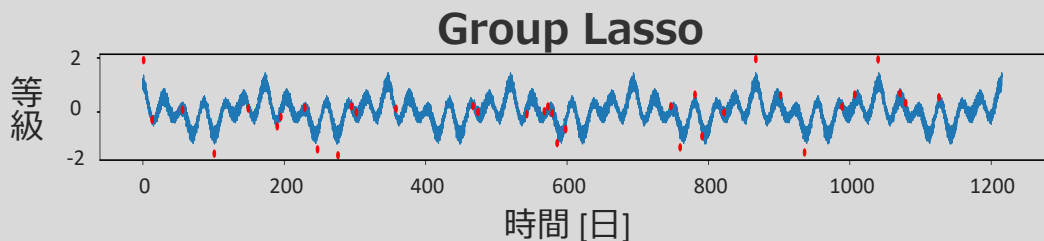
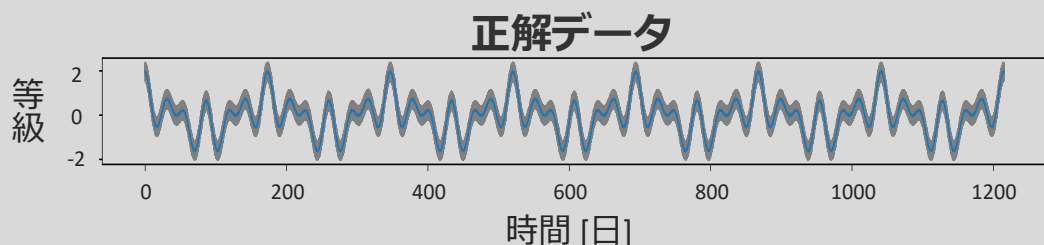
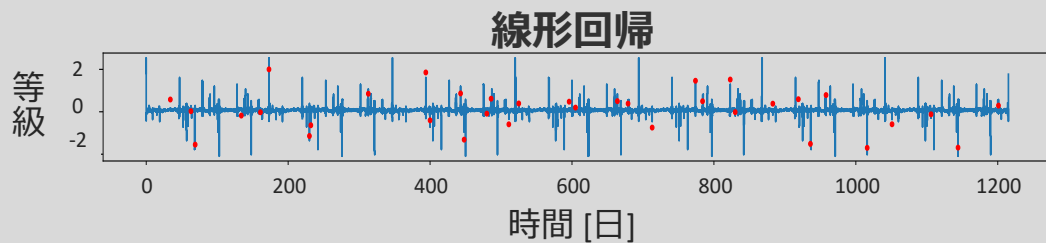
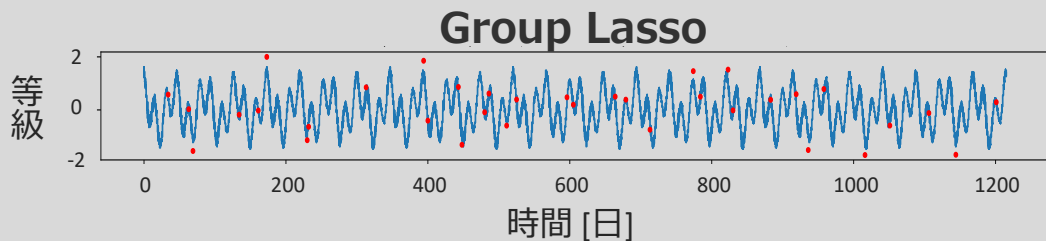
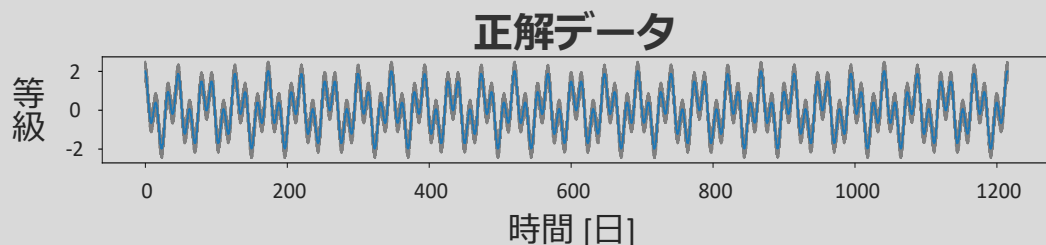
- 重要周期が不明
- 過学習の発生

周期解析結果



周期解析の堅牢性評価

その他の例



周期解析の堅牢性評価

定量的な評価

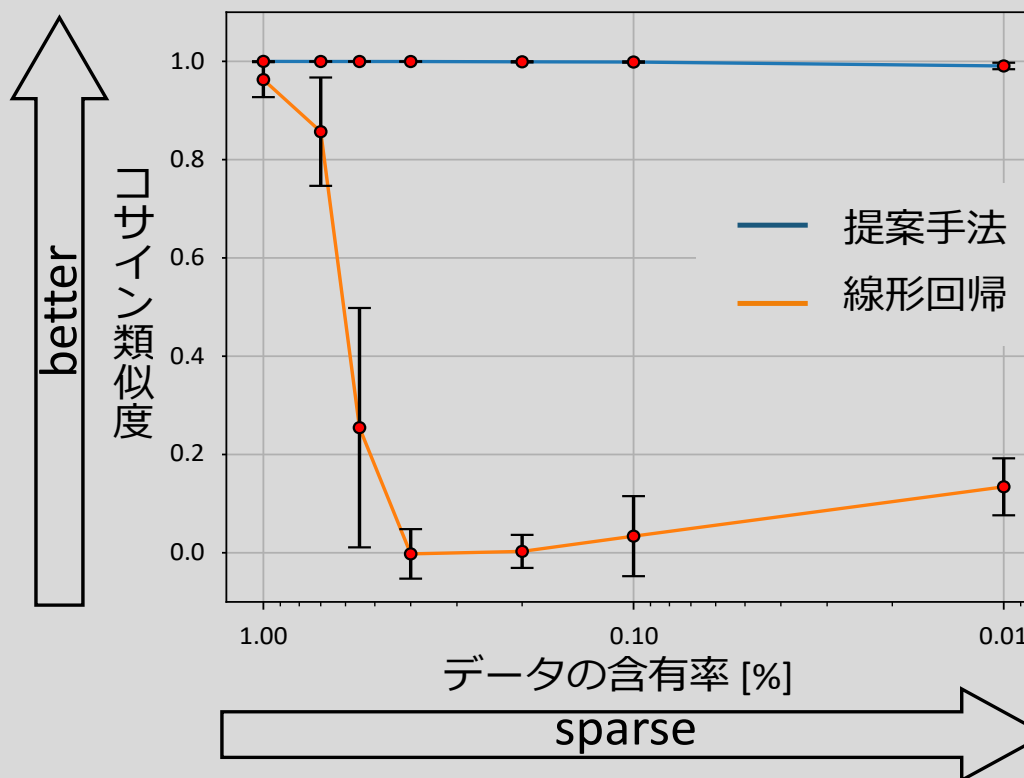
評価指標: コサイン類似度

- ベクトル間の類似度に関する尺度
- 距離ベースではなく角度ベース

結果

スパースなデータに対しても
高精度な結果を維持している

データの含有量と整合率の関係性



周期解析の堅牢性評価

定量的な評価

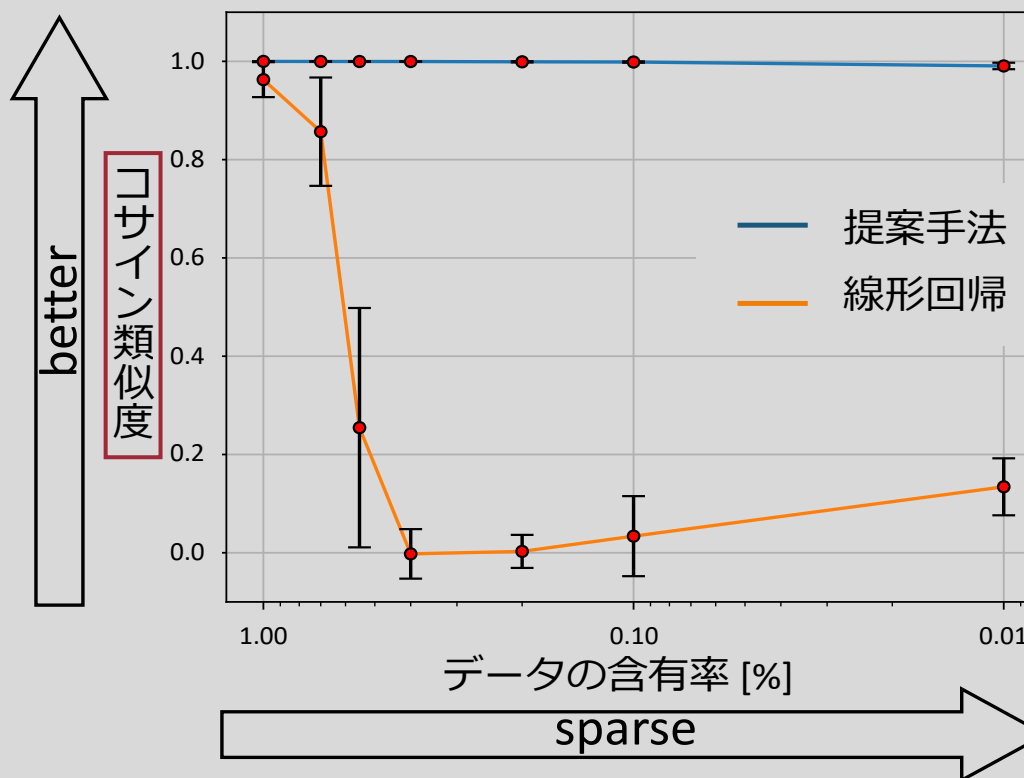
評価指標: コサイン類似度

- ベクトル間の類似度に関する尺度
- 距離ベースではなく角度ベース

結果

スパースなデータに対しても
高精度な結果を維持している

データの含有量と整合率の関係性



周期解析の堅牢性評価

定量的な評価

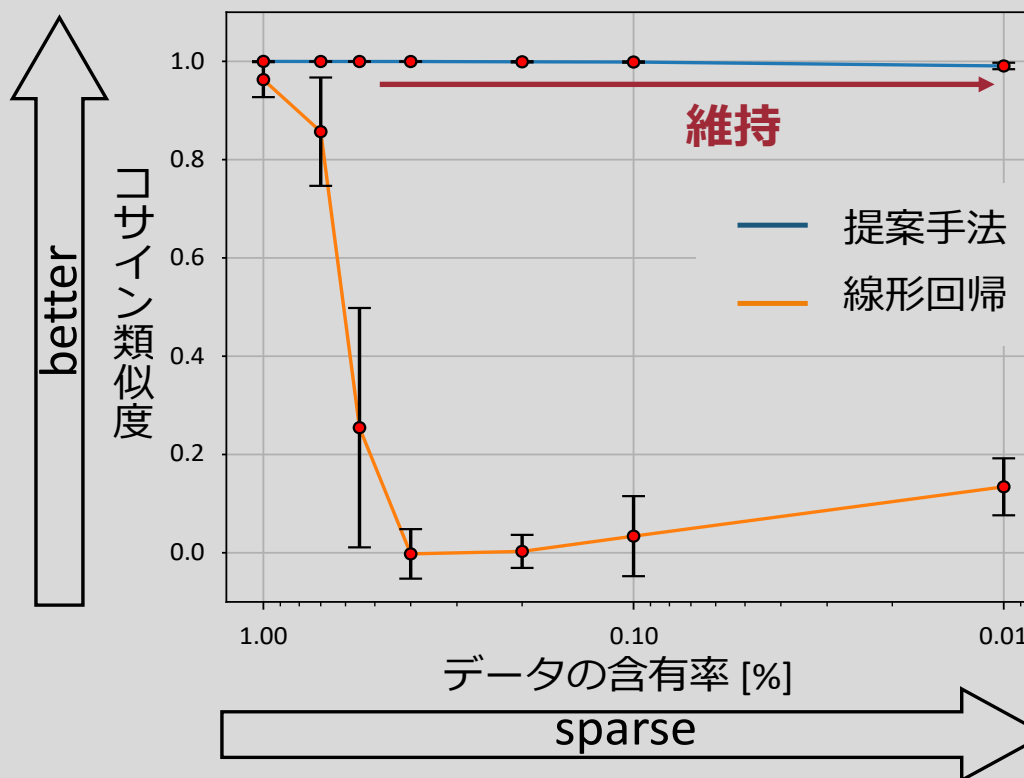
評価指標: コサイン類似度

- ベクトル間の類似度に関する尺度
- 距離ベースではなく角度ベース

結果

スパースなデータに対しても
高精度な結果を維持している

データの含有量と整合率の関係性



まとめ

目的

多数の天体データセットから**周期情報**を活用した**変動天体の検出**

課題点

天体データの**スパース性**
特徴量候補の**不足**

解決策

スパースモデリングの活用
ドメイン知識の追加

今後の課題

並列化による高速化
他のスパースモデリングの導入による提案モデルの性能向上