

2023年1月

修士論文

機械学習を用いた激変星を特徴
づける物理量の研究

M212234 古賀柚希

指導教員

植村誠 准教授

広島大学

先進理工系科学研究科

Abstract

宇宙には変光星と呼ばれる、明るさの急激な変化を示す天体がある。変光星の中には激変星と呼ばれる天体があり、種類にもよるが、数十年に一度といった頻度で大規模な爆発を起こす。この爆発に伴って、数時間～数日で急激な増光を見せるが、その突発性から、爆発直後を捉えることは難しい。それゆえ、激変星の爆発初期には未解明な物理が多く潜んでいる。この物理を探るためには、爆発直後の天体の適切な追跡観測が必要不可欠である。しかし、適切な追跡観測には専門家の知見が必要だが、専門家が常に観測場所に居続けることは現実的ではない。そこで、激変星の爆発などの突発現象に対する追跡観測の意思決定・実行を自動化したシステムがあれば有用ではないかと考え、我々はシステム Smart Kanata(以下 SmartK)の開発に取り組んでいる。Kanataは、広島大学が保有する「かなた望遠鏡」に由来している。

このシステムでは機械学習を用いて、激変星と似た振る舞いを示す天体など、合計5つの天体の型判別問題を扱う。しかし、機械学習で用いている教師データには欠損値が多く含まれるため、この欠損を上手く扱える判別モデルの検証を行う必要がある。このような欠損の扱いは、本研究のみならず他の天文学や機械学習の分野においても共通の問題である。

機械学習の分野において、判別に用いる座標や距離などの量のことを特徴量というが、激変星の判別に有効な特徴量は自明ではない。さらに、特徴量はただ多くすればいい訳ではない。機械学習においては、教師データに基づいた未知のデータに対する予測性能である汎化性能が高いことが望まれる。また、あまりにも教師データに偏った過学習といわれるような状態は避けたい。つまり、激変星の判別に特に有効な特徴量が明らかになれば、汎化性能を高めつつ過学習を避けたモデルの構築が可能となる。本研究の目的は、そのような特徴量を明らかにすると共に、激変星への理解を深めることである。

本研究では、激変星を含めた2704のサンプルを教師データとし、教師データの扱い方がそれぞれ異なる3つの判別モデルを用いて、SmartKに適していると思われる判別モデルを明らかにした。用いたモデルは、確率論や統計学で用いられるベイズの定理をシグモイド関数に置き換え、データからパラメータを推定し確率を求めるロジスティック回帰、ロジスティック回帰で過学習を防ぐスパース多クラスロジスティック回帰、ベイズの定理の尤度を正規分布で近似した生成モデルである。これらの3つのモデルにおいて、本研究の教師データにおける性能を評価し、生成モデルが最も適当なモデルだと結論づけた。さらに、生成モデルを用いて、激変星の判別に特に有効な特徴量も明らかにすることができた。例えば、激変星の中には矮新星というサブクラスがあり、さらに矮新星の中にもWZ Sge型矮新星というサブクラスがあるが、これまでこの2つの天体は、増光時の絶対等級において同程度の明るさを示すと考えられていた。しかし、激変星の判別に特に有効な特徴量を調べた結果、極大絶対等級も判別に有効である結果が得られ、WZ Sge型が通常の矮新星よりも高い光度をもつことが示唆された。この原因を究明することは、今後の研究課題である。

本研究で明らかになった特徴量を、これからのSmartKを用いた研究に上手く取り入れることで、さらに激変星についての理解が進むことが期待される。

Contents

1 序章	3
1.1 激変星	3
1.1.1 新星	3
1.1.2 矮新星	5
1.1.3 WZ Sge 型矮新星	6
1.2 Smart Kanata	7
1.3 Smart Kanata における突発現象の機械判別	10
1.4 研究目的	11
2 機械学習	12
2.1 判別モデル	12
2.1.1 ロジスティック回帰	13
2.1.2 スパース多クラスロジスティック回帰	14
2.1.3 生成モデル	15
2.2 性能評価の指標	15
2.2.1 混同行列	16
2.2.2 AUC(Area Under the Curve)	16
2.2.3 交差検証	17
3 結果: 判別モデル	19
3.1 混同行列	20
3.2 正解率	21
4 結果: 変数選択	26
4.1 新星と矮新星の判別	26
4.2 新星と WZ Sge 型星の判別	27
4.3 新星とミラ型星の判別	28
4.4 矮新星と WZ Sge 型星の判別	29
4.5 矮新星とフレア星の判別	30
5 議論	31
5.1 新星と矮新星	32
5.2 新星と WZ Sge 型星	33
5.3 新星とミラ型星	34
5.4 矮新星と WZ Sge 型星	35
5.5 矮新星とフレア星	37
6 まとめ	39
A 付録	41
A.1 混同行列	41
A.1.1 LR	41
A.1.2 SMLR	42
A.1.3 GM	43
A.2 正解率	44

1 序章

1.1 激変星

ロッシュローブを満した通常の恒星と白色矮星からなる連星系のことを激変星という。ロッシュローブとは、図 1 のように、近接連星系のポテンシャルを記述するロッシュモデルにおいて、ラグランジュの平衡点 L_1 を通るものを指す。星の半径がロッシュローブを越えると、星の表面のガスがポテンシャル面の最も低い場所 L_1 点から、相手の星のロッシュローブ内に流れ込む。

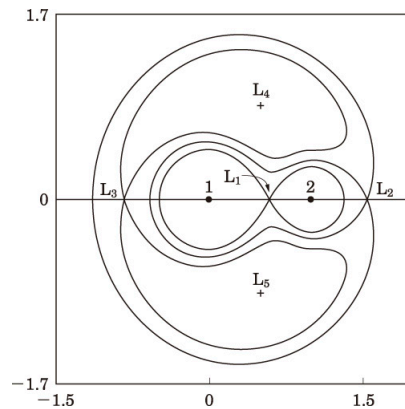


図 1: 公転軌道面上のロッシュ等ポテンシャル面。 $L_1 - L_5$ はラグランジュ点。[1]

激変星のなかには数時間から数日のタイムスケールで急激な増光を起こすものがあるが、その突発性ゆえ、観測が難しく理解されていないことが多い。また、その増光・爆発機構の違いによっていくつかの型に分類される。本研究では、新星、矮新星、WZ Sge 型矮新星について扱うため、以下で詳しく紹介する。

1.1.1 新星

恒星からのガスが白色矮星表面に降着し、水素の核燃焼が始まることで降着ガスを吹き飛ばす現象を新星爆発といい、このような現象を起こす天体を新星という。降着ガスの密度が、白色矮星表面で核融合を起こす臨界密度に達するまでには典型的には数千年～数万年かかり、このタイムスケールで爆発を繰り返す。図 2 に新星の光度曲線を示す。新星爆発を起こすと、爆発開始から数日程度で静穏時から 10 等程度増光し、数か月から数年で静穏等級に戻る。

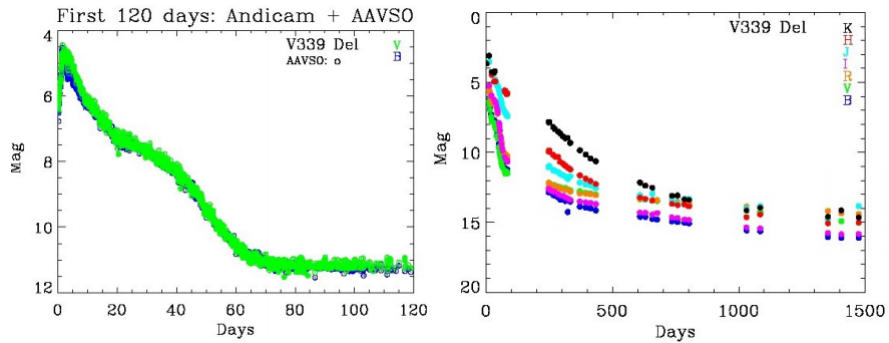


図 2: 典型的な新星の光度曲線 (V339 Del)[2]

図 3 は典型的な新星のスペクトルである。新星のスペクトルは強い水素の輝線で特徴づけられ、青方偏移した吸収線を伴うことがある。これを P Cyg プロファイルという。P Cyg プロファイルは膨張ガスの存在を示唆するものである。白色矮星表面の核融合によって降着ガスが吹き飛び、そのガスからの放射が P Cyg プロファイルを作っていると考えられている。

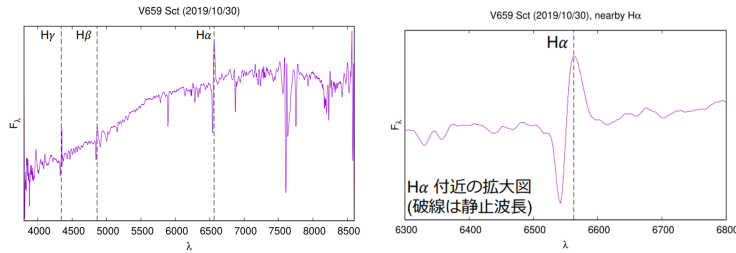


図 3: 典型的な新星スペクトル [3]

最近、新星爆発の極初期に新星としては特異なスペクトルが現れることが報告されている。図 4 に新星 T Pyx のライトカーブと極初期のスペクトルを示す。図 4 のスペクトルは、発見から 0.19 日 (4.6 時間) 後に観測されたもので、これほどの極初期に分光されたものはこの 1 例のみである。特徴として、高い電離階数の輝線が見られることが挙げられる。これらの高励起輝線は極大付近では消えたため、増光中の白色矮星表面が見えているのではないかと考えられるが、そのような新星爆発の極初期の物理状態はまだよく理解されていない。また、P Cyg プロファイルは見られないが、これは膨張ガスが生まれる前を見ているからだと考えられる。

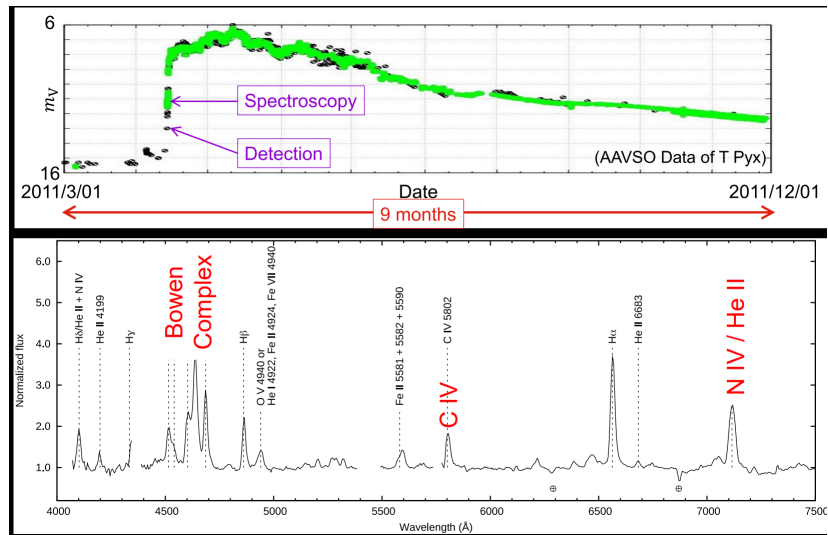


図 4: 新星 T Pyx の光度曲線 (上) と極初期スペクトル (下)[4]

このように、新星の爆発直後は未知の、理解が進んでいない状態が存在する。この状態を研究するためには新星爆発の発見後、即座に分光観測を行うことが重要となる。

1.1.2 矮新星

恒星からのガスが白色矮星表面に降着する際、すぐには降着せず、円盤外縁部に溜めこまれる。溜めこまれた質量がある臨界値を越えると、円盤が不安定になり、それまでに溜めこまれた質量が高い降着率で白色矮星へ降着し、降着円盤が増光する。これを矮新星アウトバーストといい、このような現象を起こす天体を矮新星という。図5に矮新星の想像図、光度変化の模式図を示す。矮新星アウトバーストは新星爆発と異なり、降着円盤にガスが溜まるまでが典型的には数 100 日程度なので、このタイムスケールで数等程度の増光を繰り返す。

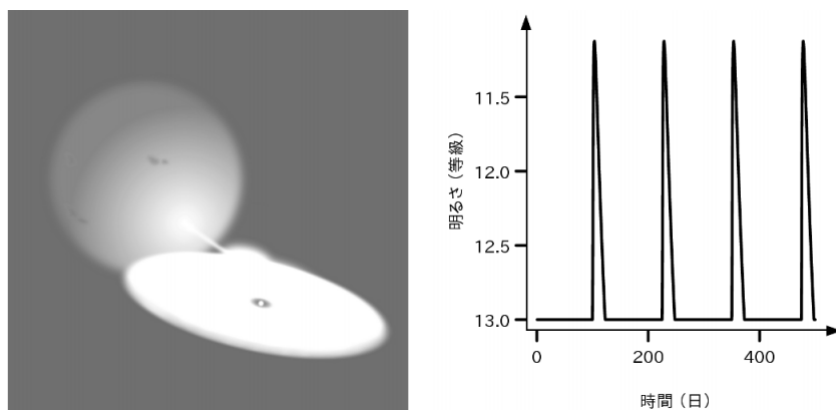


図 5: 矮新星の想像図 (左) と光度変化の模式図 (右)[5]

1.1.3 WZ Sge 型矮新星

矮新星は降着円盤の安定性などからさらにいくつかの型に分類される。通常光度より 2~4 等程度明るくなるノーマルアウトバーストを 10 日~数 100 日の間隔で繰り返す「SS Cyg 型」、ノーマルアウトバーストに加えてより明るく長く継続するスーパーアウトバーストを示す「SU UMa 型」、SU UMa 型の中でもスーパーアウトバーストの間隔が極端に長い「WZ Sge 型」などがある。特に、WZ Sge 型のスーパーアウトバーストの極初期には、「早期スーパーハンプ」という現象が見られる。図 6 に早期スーパーハンプの光度曲線を示す。早期スーパーハンプとは、明るさや温度が連星の公転周期 (約 80 分) で振動する現象のことで、円盤の幾何学構造を探る手掛かりになると考えられている。図 7 に、図 6 のデータから推定された降着円盤の高さ構造を示す。しかし、早期スーパーハンプはアウトバーストの最初の数日しか現れないため観測が難しく、矮新星アウトバーストの初期における円盤構造の変化はいまだによく理解されていない。この研究のためには、WZ Sge 型矮新星のアウトバーストのごく初期に、多色の連続撮像観測を行うことが重要である。

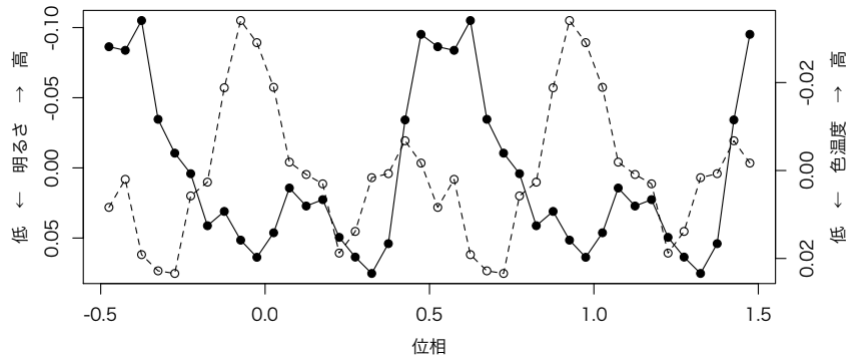


図 6: 早期スーパーハンプを表す明るさと温度の振動 [5]

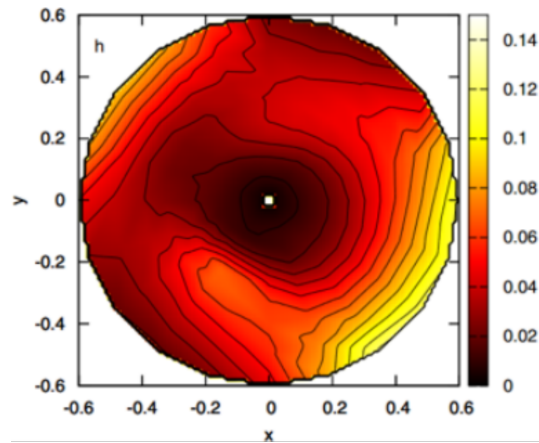


図 7: 図 6 のデータから推定された降着円盤の高さ構造 [6]

1.2 Smart Kanata

1.1 節で述べたように、新星・矮新星ともに、爆発初期のタイムスケールの短い現象の物理に未解明のことが多い。つまり、適切な追跡観測、例えば、新星の場合は発見直後の分光観測、矮新星の場合は連続測光観測、が必要不可欠となる。近年、広い視野をもつ望遠鏡による全天サーベイ観測が盛んで、機械学習の応用もあり、毎日多くの突発現象が自動的に検出されるようになっている。全天サーベイ観測プロジェクトとして、ASAS-SN(All-Sky Automated Survey for SNe) や ZTF(Zwicky Transient Facility) などがある。ASAS-SN は、主に超新星の発見を目的とした、オハイオ州立大学を中心とする研究グループのプロジェクトで、口径 14-cm といった小型の望遠鏡を南アフリカ、チリ、テキサス、ハワイに設置し

ている。ZTFは、主にカリフォルニア工科大学などが参加するプロジェクトで、カリフォルニアのパロマー天文台にある1.2-m望遠鏡を使っている。これらのプロジェクトが、突発現象を発見した際にデータを共有するTNS(Transient Name Server)というデータベースがあり、そこに毎日多くの突発現象が報告されている。

しかし、新天体の発見時は天体の型の不確定度が高く、追跡観測(多色撮像、分光、偏光、連続観測など)の手段を適切に判断する必要がある。また、発見される突発天体の全てが新星やWZ Sge型矮新星ではなく、他の様々な型の現象も報告され、なかには激変星と混同されやすい型のものもある。例えば、発見時に新星や矮新星と混同されやすい変光星型としてはミラ型変光星やフレア星が挙げられる。ミラ型変光星とは、脈動変光星と呼ばれる、星自身が膨張と収縮を繰り返すことで変光を繰り返す天体の一種で、変光の周期は約80日から約1000日である。光度の振幅が2等から10等程度と大きいので、未知のミラ型変光星の極大時が突発現象として検出されることがある。図8にミラ型変光星LX Cygの光度曲線を示す。LX Cygのように極小期まで観測できれば矮新星とは区別できるが、暗いミラ型変光星で1年から数年に1回、極大期のみが検出される場合は矮新星と混同される。

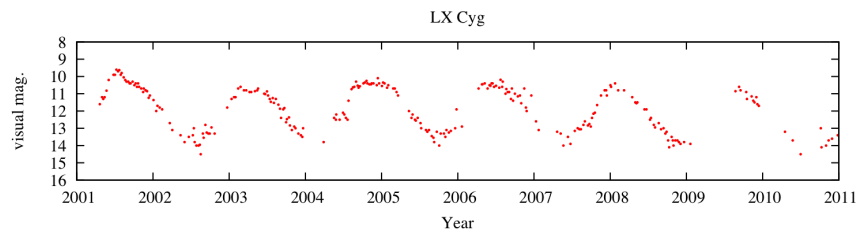


図 8: LX Cyg の光度曲線 [7]

フレア星とは、恒星やその周囲の降着円盤での磁場活動によるエネルギー開放でフレアを起こす天体であり、特に「UV Cet 型」と呼ばれるフレア星の変光タイムスケールは数秒から数時間である。図9にUV Cetの光度曲線を示す。変光のタイムスケールが激変星とは異なるので、光度曲線全体が観測されれば区別できるが、フレア中の1点のみだと混同される。

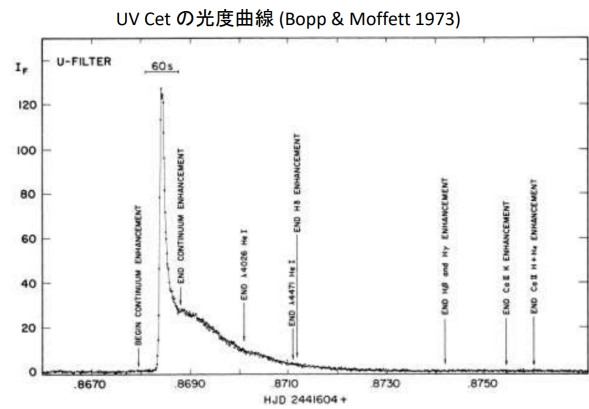


図 9: UV Cet の光度曲線 [8]

これらの変光星は適切な追跡観測を行うことで新星や矮新星と区別することができる。従来は、天体の特徴や必要な観測に熟知した観測者が、その場で追跡観測の可否を判断してきた。しかし、突発現象が発見された時に、そのような熟練した観測者が不在の時は追跡観測を行うことができず、天体の貴重な状態を見逃すことも多かった。そこで、情報理論の枠組みを用いて、この追跡観測の意思決定を自動化する研究が始まっている [9]。そのシステムの構築段階で、広島大学が所有する 1.5-m 望遠鏡の「かなた望遠鏡 (図 10)」を使っていることから、このプロジェクトを Smart Kanata(以下、SmartK) と呼んでいる。



図 10: 広島大学が所有するかなた望遠鏡 [10]

SmartK の全体像を説明する。まず、突発現象の報告がされている TNS などのデータベースを常に見張り、突発現象の報告を受ける。新天体の報告があった場合、座標、距離、静穏時対応天体など、新天体の情報をオンラインデータベー

スから収集する。次に、事前に用意してあるカタログから、新天体の情報がある特徴量が揃っているサンプルのみを抽出し、それらを教師データとして機械判別し、天体 k である確率 $p(k)$ を得る。SmartK において、 $k =$ 新星, 矮新星, WZ Sge 型星, ミラ型星, フレア星の 5 つの型を想定している。その次の段階として、SmartK では情報エントロピーという、確率から求められる不確実性を表す量を用いているのだが、これを指標として追跡観測の判断を行う。最後に時刻、天体位置、天気などのステータスを確認し、観測可能であれば望遠鏡を駆動させる。以上が、SmartK 全体の流れである。

1.3 Smart Kanata における突発現象の機械判別

ここでは、機械判別で確率 $p(k)$ を求める詳細を述べる。用いている教師データについて、図 11 にデータテーブルを抜粋したものを示す。1 番上の行に書かれてあるのは説明変数、機械学習の言葉で言えば特徴量で、SmartK の機械判別で用いている特徴量の一覧を表 1 に示す。

図 11: 本研究で用いた教師データのデータテーブル (一部抜粋)

記号・略称	意味
l, b	銀河座標における経度・緯度
d_{kpc}	天体までの距離
gal_abs_z	銀河面からの距離
$AbsMag_out$	アウトバースト時の絶対等級
$AbsMag_qui$	静穏時の絶対等級
$AbsMag_J$	J バンドにおける絶対等級
$Ampl$	静穏時とアウトバースト時の等級の差である振幅
$g - r, r - i, i - z, j - h, h - k, i - k$	可視光や近赤外線の色

表 1: 本研究で用いた教師データにおける特徴量の詳細

この教師データでは、一般的な変光星のカタログである AAVSO International

Variable Star Index VSX から、新星、矮新星、WZ Sge 型星、ミラ型星、フレア星の位置と極大等級を得て、これをサンプルとしている。サンプルのその他の特徴量は、距離は Gaia DR2 カタログ、静穏時の可視光等級は PanSTARRS DR1 カタログ、なければ SDSS DR12 カタログ、近赤外線等級は 2MASS カタログから得ている。図 11 の 2 行目以下の 1 行ずつがサンプル 1 つ 1 つで、これら 14 個の特徴量を要素にもつベクトル x を特徴量ベクトルと呼ぶ。また、サンプル数はそれぞれ Nova が 298、DN が 1041、WZ が 165、Mira が 779、UV が 421 の合計 2704 である。

図 11 には数値以外に NaN となっている部分があるが、これは欠損値である。欠損値は、天体までの距離が分からない、天体が暗いなどの理由で生じる。通常の機械判別のモデルでは欠損は想定しておらず、少量の欠損であれば欠損値を推定することもできるが、この教師データには、このような欠損値が全体の 37% ある。この欠損値をどのように扱うのが本論文の大きな命題となっているが、欠損値の扱いは、本研究のみならず他の天文学や機械学習の分野においても共通の問題である。

さらに、表 1 の 14 特徴量全てを使った判別モデルが最適とも限らず、過適合なモデルになる可能性があるため、天体の判別に有効な特徴量だけで判別モデルを構築したいが、そのような特徴量は自明ではない。14 個の特徴量のなかから、天体の判別に有効な特徴量が明らかになれば、モデルの汎化性能の向上だけでなく、天体の理解も深めることができる。このように、多くの特徴量の中から必要な特徴量をデータから客観的に選択し、選ばれた特徴量から対象に関する新たな知見を得るアプローチはデータ駆動型科学と呼ばれる。

1.4 研究目的

本研究の目的は、突発現象を判別するために有効な特徴量を明らかにし、その特徴量から新星や矮新星に関する新たな知見を得ることである。また、それを調べるために、欠損の多い教師データに適した判別モデルも調べる。

以降、第 2 章では機械学習について、本研究で扱うモデルや性能評価の指標について述べる。第 3 章ではモデルごとの性能を検証結果を踏まえて比較し、本研究で扱う教師データに最適な判別モデルを決める。第 4 章ではある 2 つの天体型について、14 特徴量の全ての組み合わせで機械判別した結果を示し、第 5 章ではその結果について議論する。第 6 章で、本研究のまとめを述べる。

2 機械学習

機械学習とは、人工知能を実現するための技術領域の1つで、コンピュータを使って大量のデータを学習することでパターンやルールを見つけ出し、人間がしている学習にあたる仕組みを実現する技術や手法の総称である。つまり、色々なことを教えなくても勝手に学習する能力をコンピュータに与えることを指す。また、機械学習は「教師あり学習」「教師なし学習」「強化学習」に分類されるが、本研究では学習データとして正解を与える「教師あり学習」を採用している。

2.1 判別モデル

教師あり学習には「分類」と「回帰」がある。「分類」の主な目的は、データが属するクラスを予測することである。予測するクラス数が2クラスの場合は2値分類、2クラスより多い分類予測については多クラス分類と呼ばれる。「回帰」の主な目的は、連続値などの値の予測をすることである。具体例としては、広告予算の増加による商品の売り上げの増加を予測することなどがある。本研究で我々が行うのは、「分類」の中でも特に多クラス分類である。今回の問題では天体の型を5つ想定しているので、5クラスのカテゴリ問題となる。この5クラスのそれぞれに属する確率を求めるのだが、まずはその前提となるベイズの定理について紹介する。

ベイズの定理とは、「事前確率が尤度を受けて、どう変化するかを示す事後確率を求めるための方法」である。事前確率は元々持っている信念や考え、尤度は新しいデータや経験のことである。これを数式で表すと以下ようになる。

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)}$$

ここでは、 H を仮説、 D をデータと読むと分かりやすい[11]。すると、それぞれの文字の意味合いは以下ようになる。

- 事前確率 $p(H)$: 事前にデータ D が何もない状態で、仮説 H が正しい確率
- 尤度 $p(D|H)$: 仮説 H が正しいとした時の、データ D の尤もらしさ
- 周辺尤度 $p(D)$: 仮説 H の真偽を問わず、データ D を観察する可能性
- 事後確率 $p(H|D)$: データ D を踏まえて、仮説 H が正しい確率

これを今回の我々の問題に落とし込むために、仮説 H を「クラス C_k に属する」、データ D を「特徴量ベクトル \mathbf{x} 」と読み替える。

$$p(C_k|\mathbf{x}) = \frac{p(\mathbf{x}|C_k)p(C_k)}{p(\mathbf{x})}$$

- 事前確率 $p(C_k)$: 前に特徴量ベクトル \mathbf{x} が何もない状態で、クラス C_k に属する確率

- 尤度 $p(\mathbf{x}|C_k)$: クラス C_k に属するとした時の、特徴量ベクトル \mathbf{x} の尤もらしさ
- 周辺尤度 $p(\mathbf{x})$: クラス C_k を問わず、特徴量ベクトル \mathbf{x} を観察する可能性
- 事後確率 $p(C_k|\mathbf{x})$: 特徴量ベクトル \mathbf{x} を踏まえて、クラス C_k に属する確率

今回我々が求めるのは事後確率 $p(C_k|\mathbf{x})$ である。本研究では、欠損値の扱い方や決定境界の引き方が異なる3つの判別モデルを用いて、それぞれの性能を比較した。以下ではその3つの判別モデルを紹介する。

2.1.1 ロジスティック回帰

ベイズの定理を以下のように変形する。

$$p(C_k|\mathbf{x}) = \frac{p(\mathbf{x}|C_k)p(C_k)}{\sum_i p(\mathbf{x}|C_i)p(C_i)} = \frac{\exp(a_k)}{\sum_i \exp(a_i)}$$

ここで、 $a_k = \log(p(\mathbf{x}|C_k)p(C_k))$ である。さらに、 a_k を \mathbf{x} の線形結合で以下のように表せるとする。

$$a_k = \mathbf{w}_k^T \mathbf{x}$$

このような変形は、データが生成される確率分布が正規分布である時に厳密に正しいことが知られている。このモデルパラメータ \mathbf{w}_k をデータから最尤推定して構築する判別モデルをロジスティック回帰 (Logistic Regression : LR) という。

最尤推定について説明する。教師データにおいて、 i 番目のサンプルがクラス C_k に属する確率 $p(C_k|\mathbf{x}_i)$ と、 j 番目のサンプルがクラス C_k に属する確率 $p(C_k|\mathbf{x}_j)$ は明らかに独立である。このような場合、両方が同時に成り立つ確率は、以下のようにそれぞれの確率を掛けて表すことができ、これを同時確率と呼ぶ。

$$p(C_k|\mathbf{x}_i, \mathbf{x}_j) = p(C_k|\mathbf{x}_i)p(C_k|\mathbf{x}_j)$$

また、教師データについては、1つ1つのサンプルの型は明らかである。これを、新たな目的変数 $\mathbf{y}_i (1 \leq i \leq N)$ を導入して表現する。 N はサンプル数で、SmartK の教師データでは $N = 2704$ である。 \mathbf{y}_i はクラス数 K 次元ベクトル、つまり5次元ベクトルで、 $\mathbf{y}_i = (1, 0, 0, 0, 0)$ のように書く。この場合、 i 番目のサンプルが属するクラスは新星であるというように読む。 $\mathbf{y}_i = (0, 1, 0, 0, 0)$ なら矮新星、 $\mathbf{y}_i = (0, 0, 1, 0, 0)$ なら WZ Sge 型星、 $\mathbf{y}_i = (0, 0, 0, 1, 0)$ ならミラ型星、 $\mathbf{y}_i = (0, 0, 0, 0, 1)$ ならフレア星と表現する。これを用いると、上の同時確率の式は以下のように書き替えることができる。

$$p(y_{ik}, y_{jk}|\mathbf{W}) = \prod_{i'=i,j} \prod_{k'} \left\{ \frac{\exp(\mathbf{w}_{k'}^T \mathbf{x}_{i'})}{\sum_{j'} \exp(\mathbf{w}_{j'}^T \mathbf{x}_{i'})} \right\}^{y_{i'k'}}$$

ここで、 y_{ik} は \mathbf{y}_i の k 個目の要素、 \mathbf{W} は5つの \mathbf{w}_k をまとめて書いたものである。 N 個のサンプルが生成される同時確率 $p(\mathbf{Y}|\mathbf{W})$ は、 N 個のサンプルがクラ

ス C_k に属する確率を全て掛け算することで、以下のように与えられる。

$$p(\mathbf{Y}|\mathbf{W}) = \prod_{i'} \prod_{k'} \left\{ \frac{\exp(\mathbf{w}_{k'}^T \mathbf{x}_{i'})}{\sum_{j'} \exp(\mathbf{w}_{j'}^T \mathbf{x}_{i'})} \right\}^{y_{i'k'}}$$

このように、事後確率 $p(C_k|\mathbf{x})$ はパラメータ \mathbf{W} によって決まる \mathbf{Y} の尤度 $p(\mathbf{Y}|\mathbf{W})$ とみなせる。データからモデルパラメータを推定したい時、この尤度 $p(\mathbf{Y}|\mathbf{W})$ が最も高くなるように \mathbf{W} を推定する方法を最尤法もしくは最尤推定という。このように、パラメータをもつモデルをデータに当てはめる時、何かしらの関数を最小化、または最大化するようにパラメータを探すが、そのような関数は目的関数と呼ばれる。ロジスティック回帰では、目的関数 $E(\mathbf{W})$ は以下のように設定される。

$$E(\mathbf{W}) = -\log(p(\mathbf{Y}|\mathbf{W})) = -\sum_{i'} \sum_{k'} y_{i'k'} \log \frac{\exp(\mathbf{w}_{k'}^T \mathbf{x}_{i'})}{\sum_{j'} \exp(\mathbf{w}_{j'}^T \mathbf{x}_{i'})}$$

最大にしたい同時確率 $p(\mathbf{Y}|\mathbf{W})$ の負の対数をとったので、これを最小にする解を $\hat{\mathbf{W}}$ と書き、以下のように表される。

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W}} E(\mathbf{W})$$

左辺の $\hat{\mathbf{W}}$ は \mathbf{W} の推定値を意味し、右辺の $\arg \min$ は関数が最小になる点を表す。つまり、この式は「目的関数 $E(\mathbf{W})$ を最小にする \mathbf{W} をその推定量とする」という意味である。ロジスティック回帰は特徴量 \mathbf{x} の線型結合 $\mathbf{w}_k^T \mathbf{x}$ を考える一般化線形モデルであるため、判別モデルの決定境界は特徴量の単純な線形関数になる。

また、LR が教師データとして扱えるサンプルは、判別したいサンプルと同じ特徴量が揃っているサンプルのみである。本研究で用いている教師データには欠損値が多く含まれるため、特徴量が増えるほど、教師データとして扱えるサンプル数が減ることになる。この特徴は、2.1.2 節で紹介する SMLR にも共通する。

2.1.2 スパース多クラスロジスティック回帰

2.1.1 節のロジスティック回帰では、単に目的関数を $E(\mathbf{W})$ として最尤推定を行った。ここで、目的関数を以下のように置き換える。これを目的関数として構築する判別モデルをスパース多クラスロジスティック回帰 (Sparse Multinomial Logistic Regression : SMLR) という。

$$E'(\mathbf{W}) = E(\mathbf{W}) + \lambda \|\mathbf{W}\|_1$$

ここで、 $\|\mathbf{W}\|_1 = \sum_i \sum_j |w_{ij}|$ で正則化項と呼ばれ、係数 λ は正則化係数と呼ばれる。正則化とは、目的関数に別の制約を課して最小化し、モデルがデータに過剰に適合する過適合を避ける手法のことである。過適合は一見悪いことではないように思えるかもしれないが、機械学習においては未知のデータに対する予測性能、機械学習の言葉では汎化性能が重要となるので、モデルの汎化性能を下げてしまう過適合は避けなければならない。 $\|\mathbf{W}\|_1$ は \mathbf{W} の 1 次ノルムといい、1 次

ノルムを用いた正則化では、ゼロの要素が多いスパースなベクトルが推定される。こうすることで過剰な特徴量は排除され、過適合を避けるようなモデルを構築することができる。正則化係数 λ については、様々な λ について交差検証を行い、汎化性能を比較しながら決めていく。 λ が大きすぎると、データから外れた汎化性能の低いモデルになり、 λ が小さすぎると過適合なモデルになる。交差検証については、2.3 節で詳しく説明する。

また、判別モデルとは独立した手法として、SMLR ではデータに対してカーネル化という処理を施した。これを行うことで、調節すべきパラメータは増えるが非線形判別境界を扱うことができるようになる。カーネル化はモデルに依存せず、他の判別モデルや回帰モデルも含めて様々な場面で登場する重要な概念である。

2.1.3 生成モデル

ベイズの定理より、尤度 $p(\mathbf{x}|C_k)$ と事前確率 $p(C_k)$ が分かれば、特徴量ベクトル \mathbf{x} がクラス C_k に属する確率、つまり事後確率 $p(C_k|\mathbf{x})$ が求められる。このような、データが生成される過程を表したモデルを生成モデル (Generative Model : GM) という。本研究の生成モデルでは、事前確率 $p(C_k)$ は 5 つの型に均等とした。つまり $p(C_k) = 1/5$ である。尤度 $p(\mathbf{x}|C_k)$ には多変量正規分布を仮定した。多変量正規分布は以下で表される。

$$p(\mathbf{x}|C_k) = \frac{1}{\sqrt{(2\pi)^M |\Sigma|}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{u}_{C_k})^T \Sigma^{-1} (\mathbf{x} - \mathbf{u}_{C_k})\right\}$$

M は特徴量ベクトル \mathbf{x} の次元、 Σ は分散・共分散行列、 \mathbf{u}_{C_k} はクラス C_k の特徴量の平均ベクトルである。つまり、特徴量ごとの平均と分散、特徴量の全ての組み合わせの共分散を教師データから計算する。しかし、尤度は一般的には正規分布よりも複雑な分布をもつことに注意しなければならない。

LR と SMLR では、特徴量が揃っているサンプルしか教師データとして扱えないという特徴があったが、GM は全てのサンプルを教師データとして扱える。それには、少し工夫を加える必要がある。正規分布では、分散・共分散行列が分布を特徴づけるのだが、欠損値の影響で正しく計算されないことがある。そうになると、正規分布の式に含まれる行列式の値が負になる場合があり、尤度自体が計算不能になる。そこで、以下のように行列を変形する。

$$\Sigma' = \Sigma + \alpha I$$

α は 0 から 1 の範囲の実数パラメータで、 I は単位行列である。分散・共分散行列は対角成分が分散、それ以外の要素が共分散で構成されるのだが、このように変形することで、分散を共分散に対して相対的に大きくすることができる。 $|\Sigma|$ が負の値になる場合は、 $|\Sigma'|$ が負の値にならず、且つデータから外れすぎないように、できるだけ小さい α の値を選ぶ。

2.2 性能評価の指標

以下ではモデルの性能評価に用いた指標を、2 クラス判別を例に紹介する。

2.2.1 混同行列

	C_A (真)	C_B (真)
C_A (推)	18(TP)	2(FP)
C_B (推)	2(FN)	18(TN)

表 2: 混同行列の例

表 2 は混同行列と呼ばれるもので、対角成分が推定に成功したサンプル数で、非対角成分が推定に失敗したサンプル数である。例えば、2 行 1 列成分は、実際はクラス C_A に属するが誤ってクラス C_B に属すると推定したサンプルが 2 つ、ということを表す。混同行列は多クラス判別の場合にも拡張できるので、それぞれの混同行列を見ることで、どのような特徴量が使われている時に、どの天体の判別が得意か、もしくはどの天体の判別を間違えやすいのか、ということが調べられる。

2.2.2 AUC(Area Under the Curve)

表 2 において、以降では推定結果が C_A なら陽性、 C_B なら陰性と呼ぶことにする。表 2 の混同行列の 4 つの領域について、それぞれ以下のように呼ばれる。

- 真陽性 (True Positive: TP): 1 行 1 列成分、推定結果が C_A で実際のクラスも C_A
- 偽陽性 (False Positive: FP): 1 行 2 列成分、推定結果が C_A で実際のクラスは C_B
- 真陰性 (True Negative: TN): 2 行 2 列成分、推定結果が C_B で実際のクラスも C_B
- 偽陰性 (False Negative: FN): 2 行 1 列成分、推定結果が C_B で実際のクラスは C_A

モデルの性能を調べるなら、最も単純には正解率がある。これは、モデルによって正しく判別された教師データの割合、つまり $(TP+TN)/(TP+FP+TN+FN)$ である。表 2 の場合は $(18+18)/(18+2+18+2) = 0.9$ となる。正解率は多クラス判別の場合にも用いやすく分かりやすい指標だが、本研究で扱う教師データのように、教師データのサンプル数が天体の型によって偏りがある場合は、必ずしもモデルの性能を正しく評価できているとは言えない。例えば、クラス C_A に属するサンプルが 10^4 個、クラス C_B に属するサンプルが 10 個の教師データに対して、もし全てのサンプルをクラス C_A に判別するモデルだったとしても、正解率を計算すると $(10^4 + 0)/(10^4 + 10) \sim 0.999$ となり、高い性能だと判断してしまう。そこで、本研究ではサンプル数の偏りを考慮した評価指標を用いている。それを紹介する。

まず、いくつかの言葉を定義する。 $TP/(TP + FN)$ で表される量を真陽性率 (True Positive Rate: TPR) といい、クラス C_A に属するサンプルが正しく C_A に判別された割合である。逆に、 $FP/(FP + TN)$ で表される量を偽陽性率 (False Positive Rate: FPR) といい、クラス C_B に属するサンプルが誤って C_A に判別された割合である。表 2 の場合、 $TPR = 18/(18+2) = 0.9$ 、 $FPR = 2/(2+18) = 0.1$ となる。TPR が 1 に近い値ならモデルの性能が良い、FPR が 1 に近い値ならモデルの性能が悪いと判断できるので、この 2 つを考慮した指標が必要になる。 $TPR = 1$ 、 $FPR = 0$ の時が理想的なモデルといえる。

また、ここまで「サンプルがクラス C_A (もしくはクラス C_B) に判別された」と言ってきたが、その意味を確認しておく。判別モデルの出力は $p(C_A|x_i)$ 、つまり確率という形なので、「 $p(C_A|x_i) > P$ ならサンプル x_i はクラス C_A に属する」と判断するための決定境界 P を設定しなければならない。つまり、決定境界 P の設定の仕方、混同行列のそれぞれの成分は変動し、TPR と FPR も変動する。これを様々な P で計算すると図 12 のような曲線を描くことができ、これを受信者操作特性 (Receiver Operating Characteristic: ROC) 曲線という。図 12 は、左と右でそれぞれ性能の異なる判別モデルの ROC を描いたものである。左右ともに、緑点で表されている部分が、 $P = 0.5$ の時の点である。この 2 つを比較すると、右の ROC 曲線の方が理想的な状態 ($TPR = 1$, $FPR = 0$) から離れていて、モデルの性能が低いことを意味する。

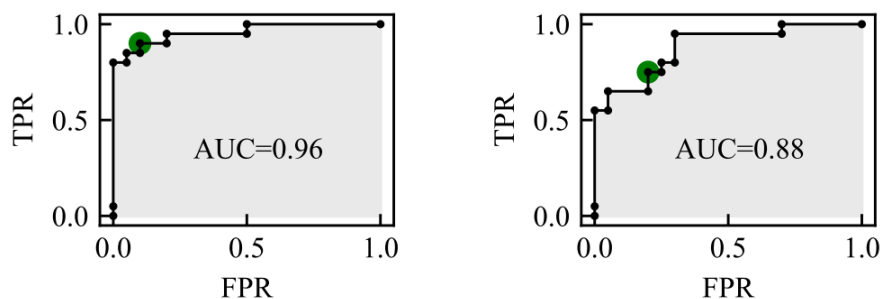


図 12: 受信者操作特性 [12] 曲線の例。緑点で表されている部分は $P = 0.5$ の時の点で、ROC 曲線より下の灰色の領域の面積が AUC を表す。

ROC 曲線を用いたモデル性能の指標として、ROC 曲線より下の領域の面積である曲線下面積 (Area Under the Curve: AUC) がある。ROC の AUC は、定義上 0 から 1 までの値をとる。図 12 で灰色になっている領域の面積が AUC で、左の方がより 1 に近い値なので高い性能だといえる。教師データのサンプル数がクラスごとに異なる状況で、AUC はモデルの性能評価の指標としてよく使われる。

2.2.3 交差検証

手元のデータを k 個に分割し、 $k - 1$ 個の訓練用データと 1 個の検証用データにランダムに分ける検証方法を k 分割交差検証法という。検証データを変えていく

ことで k 個の評価指標が得られ、その平均を取ることでモデルの汎化性能の指標が得られることが利点である。本研究では $k = N$ (サンプル数)、つまり教師データの全サンプルを 1 つずつ検証用データとして交差検証を行っている。また、判別モデルの性能評価には多クラス判別として正解率を、変数選択の性能評価にはいくつかの組み合わせの 2 クラス判別として AUC を用いている。

3 結果: 判別モデル

本研究の目的は、突発現象を判別するために有効な特徴量を明らかにすることだが、そのためにまず、欠損の多い教師データに適した判別モデルを調べなくてはならない。2.1.1~2.1.3節で紹介した3つのモデル LR,SMLR,GM は、大きく分けると以下のような特徴をもつ(表3)。

モデル	教師データ	カーネル化 (非線形境界)
LR	特徴量の揃っているサンプル	-
SMLR	特徴量の揃っているサンプル	√
GM	全てのサンプル	-

表 3: 3つのモデル毎に用いた教師データの特徴とカーネル化の有無

SmartK の新天体の判別において、判別に使える特徴量は新天体ごとに異なる。例えば、14 特徴量の全てが既知の場合は、教師データの全ての情報から判別モデルを構築できるが、座標 (l, b) と振幅 $Ampl$ しか特徴量が既知でない場合は、それら3つの特徴量のみを用いた判別モデルを教師データから構築し、新天体のクラス確率を得る。したがって、SmartK では新天体ごとに、利用できる特徴量に合わせて判別モデルを構築している。LR, SMLR, GM の性能評価は、典型的な特徴量の組み合わせごとに行う。この3つのモデルにおいて、以下のように特徴量の組み合わせを変え、まずはモデルごとの性能を比較する。

1. 座標のみ (l, b) 2704 サンプル
2. 座標、距離関係 $(l, b, d_kpc, gal_abs_z, AbsMag_out)$ 1745 サンプル
3. 座標、静穏時可視 $(l, b, Ampl, g-r, r-i, i-z)$ 1427 サンプル
4. 座標、静穏時近赤外 $(l, b, j-h, h-k)$ 1556 サンプル
5. 座標、距離関係、静穏時可視 $(l, b, d_kpc, gal_abs_z, AbsMag_out, Ampl, g-r, r-i, i-z, AbsMag_qui)$ 1045 サンプル
6. 座標、距離関係、静穏時近赤外 $(l, b, d_kpc, gal_abs_z, AbsMag_out, j-h, h-k, AbsMag_J)$ 1225 サンプル
7. 座標、静穏時可視、静穏時近赤外 $(l, b, Ampl, g-r, r-i, i-z, j-h, h-k, i-k)$ 865 サンプル
8. 14 特徴量全て 724 サンプル

これらの特徴量の組み合わせにおいて、LR,SMLR,GM の3つのモデルで交差検証を行った結果を、混同行列と正解率で示す。

3.1 混同行列

代表的なものとして、座標・距離関係・静穏時可視の10特徴量が使われた判別における、3つのモデルの混同行列を示す(表4~表6)。それぞれの混同行列の左上には正解率を示した。また、全ての組み合わせの混同行列は付録に示す。

正解率:94.16%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	48	0	0	0	0
矮新星	0	358	42	0	2
WZ Sge 型星	0	5	53	0	0
ミラ型星	0	0	0	257	2
フレア星	2	8	0	0	268

表 4: LR での判別による混同行列

正解率:93.30%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	46	1	1	0	0
矮新星	2	349	43	0	8
WZ Sge 型星	0	5	53	0	0
ミラ型星	2	0	0	256	1
フレア星	1	6	0	0	271

表 5: SMLR での判別による混同行列

正解率:88.42%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	45	2	0	1	0
矮新星	3	312	83	0	4
WZ Sge 型星	0	5	53	0	0
ミラ型星	1	0	0	257	1
フレア星	3	18	0	0	257

表 6: GM での判別による混同行列

いずれのモデルでも、混同行列の 2 行 3 列成分、つまり実際は WZ Sge 型星のサンプルを誤って矮新星だと判別してしまう例が多いことが分かる。また、他の成分を見ると、モデルによって大きな差があるということもなく、判別が得意不得意な天体型の組み合わせがあるわけではないことも分かる。これは、表 4～表 6 の座標・距離・静穏時可視以外の特徴量の組み合わせでも同様であった。

3.2 正解率

混同行列に続いて、座標・距離関係・静穏時可視の場合と、座標・静穏時近赤外の 4 特徴量の場合、14 特徴量全ての場合について、正解率の棒グラフを示す (図 13～図 15)。キャプションの括弧内の数字は、その時用いた特徴量の種類数を表す。また、全ての組み合わせの正解率のグラフは付録に示す。

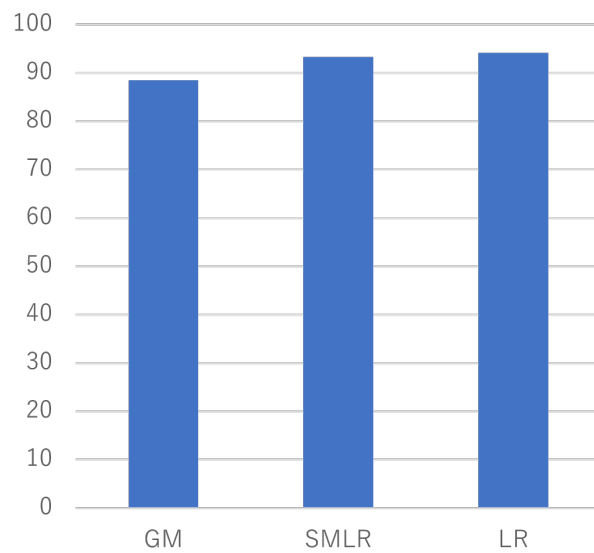


図 13: 座標、距離、静穏時可視 (10) の特徴量を用いた判別における 3 モデルの正解率

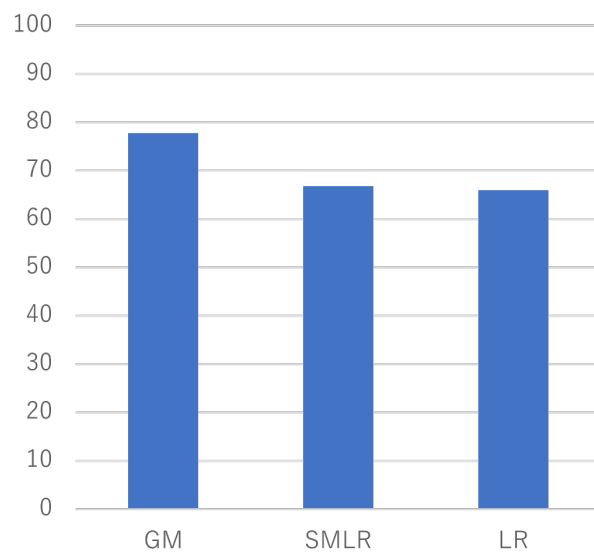


図 14: 座標、静穏時近赤外 (4) の特徴量を用いた判別における 3 モデルの正解率

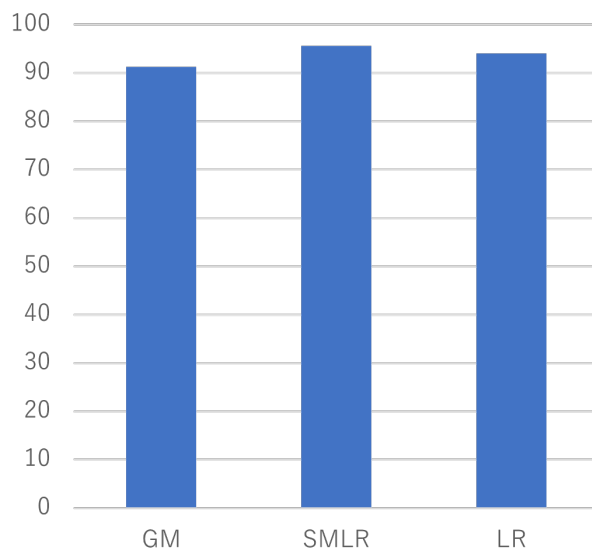


図 15: 全て (14) の特徴量を用いた判別における 3 モデルの正解率

正解率のグラフを見ると、ほとんどの場合で 3 つのモデルの間に大きな差は見られない。つまり、表 3 に示したような特徴の違いによって、判別性能に大きな差は生まれないことが分かる。

しかし、座標と静穏時の近赤外の色を用いた判別結果 (図 14) を見ると、GM の正解率が LR, SMLR と比べて 10% ほど高くなっていることが分かる。この時用いた、近赤外の色データ ($j-h, h-k$) が揃っているサンプルを表示した散布図 (図 16) を見ると、WZ Sge 型星のサンプル数が少ないことが分かる。

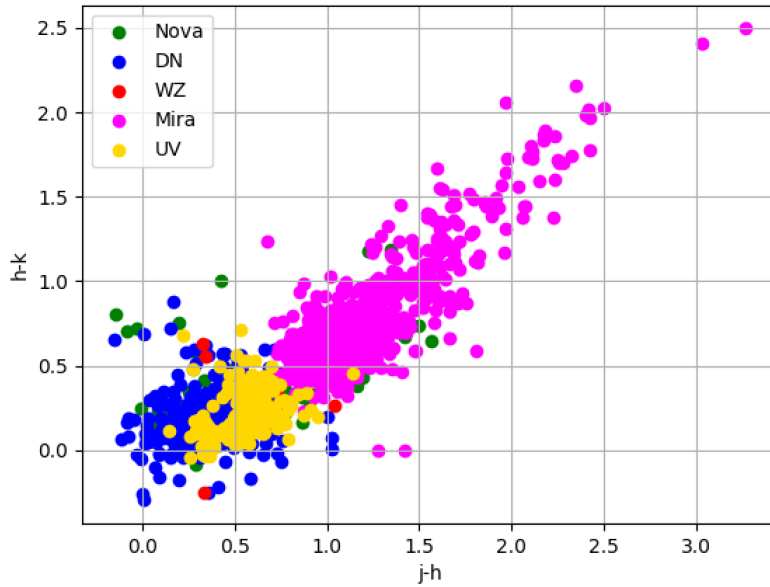


図 16: 用いた教師データにおける $j - h, h - k$ の散布図

以上のことから、本研究で用いた教師データにおいて、

- ベイズの定理の変形 (LR, SMLR) や、カーネル化という複雑な決定境界を設定 (SMLR) しなくても、正規分布を仮定した単純なモデル (GM) で同程度の性能を示す
- 座標、近赤外の色データの 4 種類の特徴量のみが利用可能な状況下においては、GM がより優れた判別性能を示す

ということが分かった。不必要に複雑なモデルは過適合になる可能性がある。また、GM では必要な特徴量が揃っていないサンプルも分散共分散の計算に使えるため、LR、SMLR に比べてより多くの情報をモデルの構築に用いている。さらに、マハラノビス距離という量を異常検知に使うことができる。マハラノビス距離の定義を以下に示す。

$$d = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

ここで、 $\boldsymbol{\Sigma}$ は分散・共分散行列、 \mathbf{x} は変数、 $\boldsymbol{\mu}$ は \mathbf{x} の平均である。マハラノビス距離は、一般的なユークリッド距離に $\boldsymbol{\Sigma}$ による補正を加えた量で、正規分布の形を考慮したものである。これによって特徴量空間での距離が与えられる。本研究の生成モデルでは、判別したいサンプルが特徴量空間のどの位置にあるのか、正規分布によって重み付けされた相対的な確率を計算する。そのため、もし想定した 5 つの型のどの型にも当てはまらない天体を判別する場合でも、5 つの型で合計 100% になるように確率が振り分けられるため、5 つの型以外の天体は検知できない。これは LR と SMLR でも同様である。一方、GM では教師データから $\boldsymbol{\Sigma}$ が

得られているため、機械判別とは別に、マハラノビス距離を計算することができる。マハラノビス距離を使えば、5つの正規分布の中心から判別したいサンプルまでの距離が分かり、どの正規分布の中心からも遠い場合は、想定した5つのどの型にも当てはまらない興味深い天体として抽出できる。これは、LR や SMLR では扱えない、GM の特徴である。

これらのことから、欠損が多い SmartK の教師データを用いた機械判別には GM が最適であると考えられる。以降、激変星の判別に効く特徴量の調査において、判別モデルは GM を採用している。

4 結果: 変数選択

この章では、特徴がより似やすい2つの天体型の2クラス判別を、14特徴量全ての組み合わせ、すなわち $2^{14} - 1 = 16383$ 通りについて行い、その天体を特徴づける物理量を調べた。具体的には、

- ほとんどの場合で判別に成功したことを示す $AUC > 0.95$
- 完全分離に成功したことを示す $AUC = 1.0$

を達成したモデルの中で、最も特徴量の数が少ないモデルを調べた。そうすることで、どんな特徴量 (物理量) がその天体の判別に有効なのかを知ることができる。

4.1 新星と矮新星の判別

新星と矮新星はどちらも激変星に分類され、静穏時の特徴は類似していると考えられる。 $AUC > 0.95$ を達成したモデルの中で、最も特徴量が少ないモデルを以下に示す。

- [*AbsMag_out*]
- [*Ampl*]

つまり、最も明るい時の極大等級 *AbsMag_out*、もしくは、爆発振幅 *Ampl* のどちらか1つの特徴量のみを用いることで、 $AUC > 0.95$ の判別モデルが構築できる。 $AUC = 1.0$ を達成したモデルの中で、最も特徴量が少ないモデルは3つの特徴量を用いたものであり、その組み合わせを表7に示す。

	<i>gal_abs_z</i>	<i>AbsMag_out</i>	<i>Ampl</i>	<i>AbsMag_qui</i>	<i>g-r</i>	<i>r-i</i>	<i>i-z</i>	<i>j-h</i>	<i>h-k</i>	<i>i-k</i>	<i>AbsMag_J</i>
パターン1	✓	✓								✓	
パターン2		✓	✓						✓		
パターン3		✓	✓							✓	
パターン4		✓	✓								✓
パターン5		✓		✓					✓		
パターン6		✓		✓							✓
パターン7		✓			✓			✓			
パターン8		✓			✓				✓		
パターン9		✓			✓					✓	
パターン10		✓			✓						✓
パターン11		✓				✓			✓		
パターン12		✓				✓					✓
パターン13		✓					✓		✓		
パターン14		✓					✓			✓	
パターン15		✓					✓				✓
パターン16		✓							✓	✓	
パターン17		✓								✓	✓
パターン18			✓	✓					✓		
パターン19			✓	✓							✓
パターン20			✓		✓						✓

表 7: 新星と矮新星の判別において AUC=1.0 を達成した特徴量の組み合わせ

これらの結果は、*AbsMag_out* と *Ampl* が判別に有効な特徴量であることを示す。また、それぞれだけでは完全分離 (AUC=1.0) 可能なモデルは作れず、そのためには他に 2 つの特徴量が必要であることがわかる。その特徴量としては近赤外線の情報 ($j-h, h-k, i-k, AbsMag_J$) が有効と言える。

4.2 新星と WZ Sge 型星の判別

WZ Sge 型星は矮新星のサブクラスであり、4.1 節と同様、静穏時の特徴量は新星と類似していると考えられる。AUC > 0.95 を達成したモデルの中で、最も特徴量が少ないモデルを以下に示す。

- [*AbsMag_out*]
- [*AbsMag_qui*]

つまり、最も明るい時の絶対等級 *AbsMag_out*、もしくは、静穏時の絶対等級 *AbsMag_qui* のどちらか 1 つの特徴量のみを用いることで、AUC > 0.95 の判別

モデルが構築できる。AUC = 1.0 を達成したモデルの中で、最も特徴量が少ないモデルは2つの特徴量を用いたものであり、その組み合わせを表8に示す。

	<i>AbsMag_out</i>	<i>AbsMag_qui</i>	<i>i - z</i>	<i>j - h</i>	<i>h - k</i>	<i>i - k</i>	<i>AbsMag_J</i>
パターン1	✓					✓	
パターン2		✓		✓			
パターン3		✓			✓		
パターン4		✓				✓	
パターン5		✓					✓
パターン6			✓				✓
パターン7						✓	✓

表 8: 新星と WZ Sge 型星の判別において AUC=1.0 を達成した特徴量の組み合わせ

これらの結果は、*AbsMag_out* と *AbsMag_qui* が判別に有効な特徴量であることを示す。また、完全分離 (AUC=1.0) するためには他に1つ特徴量が必要であることがわかる。その特徴量としては近赤外線の情報 (*j - h*, *h - k*, *i - k*, *AbsMag_J*) が有効と言える。

4.3 新星とミラ型星の判別

新星とミラ型星は、極大期のみが検出された場合は、その絶対等級がほぼ等しいため、混同されやすい。AUC > 0.95 を達成したモデルの中で、最も特徴量が少ないモデルを以下に示す。

- ["*Ampl*"]
- ["*AbsMag_qui*"]
- ["*g - r*"]
- ["*r - i*"]
- ["*i - z*"]
- ["*AbsMag_J*"]

つまり、爆発振幅 *Ampl*、静穏時の絶対等級 *AbsMag_qui*、可視光の色 *g - r*, *r - i*, *i - z*、近赤外における静穏時の絶対等級 *AbsMag_J* のいずれか1つのみを用いることで、AUC > 0.95 の判別モデルが構築できる。AUC = 1.0 を達成したモデルの中で、最も特徴量が少ないモデルは2つの特徴量を用いたものであり、その組み合わせを表9に示す。

	d_kpc	$AbsMag_out$	$Ampl$	$AbsMag_qui$	$r-i$	$i-z$	$j-h$	$h-k$	$i-k$	$AbsMag_J$
パターン1	✓					✓				
パターン2		✓			✓					
パターン3			✓		✓					
パターン4			✓			✓				
パターン5				✓	✓					
パターン6					✓		✓			
パターン7					✓			✓		
パターン8					✓				✓	
パターン9					✓					✓
パターン10						✓	✓			
パターン11						✓		✓		
パターン12						✓			✓	

表 9: 新星とミラ型星の判別において AUC=1.0 を達成した特徴量の組み合わせ

これらの結果は、 $Ampl$ 、 $AbsMag_qui$ 、 $g-r$ 、 $r-i$ 、 $i-z$ 、 $AbsMag_J$ が判別に有効な特徴量であることを示す。また、完全分離 (AUC=1.0) するためには他に 1 つ特徴量が必要であることがわかる。その特徴量としては距離 (d_kpc)、アウトバースト時の絶対等級 ($AbsMag_out$)、近赤外線の色 ($j-h$, $h-k$, $i-k$) などが有効と言える。

4.4 矮新星と WZ Sge 型星の判別

WZ Sge 型星は矮新星のサブクラスなので、類似する特徴量が多いと考えられる。この 2 クラスの判別において、1 つの特徴量で AUC > 0.95 を達成したモデルは無く、最も特徴量が少ないモデルは 2 つの特徴量を用いたものだった。それを表 10 に示す。また、この 2 クラス判別においては、AUC=1.0 を達成したモデルはなかった。

	d_kpc	gal_abs_z	$AbsMag_out$	$Ampl$	$AbsMag_qui$
パターン 1	✓			✓	
パターン 2		✓		✓	
パターン 3			✓	✓	
パターン 4			✓		✓
パターン 5				✓	✓

表 10: 矮新星と WZ Sge 型星の判別において $AUC > 0.95$ を達成した特徴量の組み合わせ

この結果は、1つの特徴量で矮新星と WZ Sge 型星を判別することは難しく、 $AUC > 0.95$ を達成するのでさえ少なくとも2つの特徴量が必要になるということを示す。ここで、 $Ampl = AbsMag_qui - AbsMag_out$ であることを考慮すると、この結果は $Ampl$ 、 $AbsMag_qui$ 、 $AbsMag_out$ のうち2つを使った時に高い判別能力をもつモデルが構築できることを意味している。

4.5 矮新星とフレア星の判別

矮新星とフレア星は静穏時の絶対等級が似ているため、フレア中の1点のみの観測だと混同される。この2クラス判別においても、4.4節の矮新星と WZ Sge 型星と同様 $AUC=1.0$ を達成したモデルはなく、さらに $AUC > 0.95$ を達成したモデルは以下の1パターンのみであった。

- [$g - r$]

この結果は、矮新星とフレア星の判別においては、可視光の色である $g - r$ が最も有効であることを示し、その他の特徴量では精度よく判別することが難しいことがわかる。

5 議論

4章の結果について議論する。まず、多くの判別において、「 $AUC > 0.95$ を達成したモデルに、近赤外線の情報 ($j-h, h-k, i-k, AbsMag_J$) を加えると $AUC=1.0$ を達成する」ことが共通点として挙げられる。4.1節の新星と矮新星の判別において、1つの特徴量で $AUC > 0.95$ を達成した $AbsMag_{out}, Ampl$ に、 $i-k$ を加えて $AUC=1.0$ を達成した例を見てみる。以下の図17,18は、 $AbsMag_{out}$ のみを用いて判別し、成功したサンプルを丸印、失敗したサンプルを星印で示し、縦軸にそれぞれ $Ampl, i-k$ を設定した散布図である。

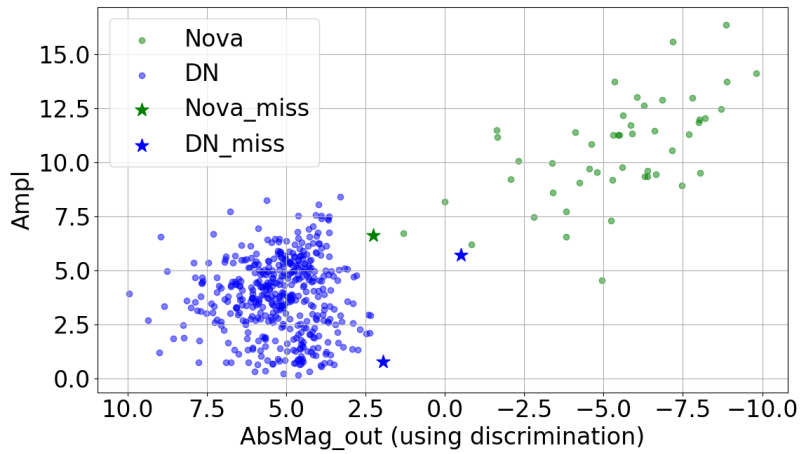


図 17: $AbsMag_{out}$ による判別結果を含めた $AbsMag_{out}, Ampl$ の散布図

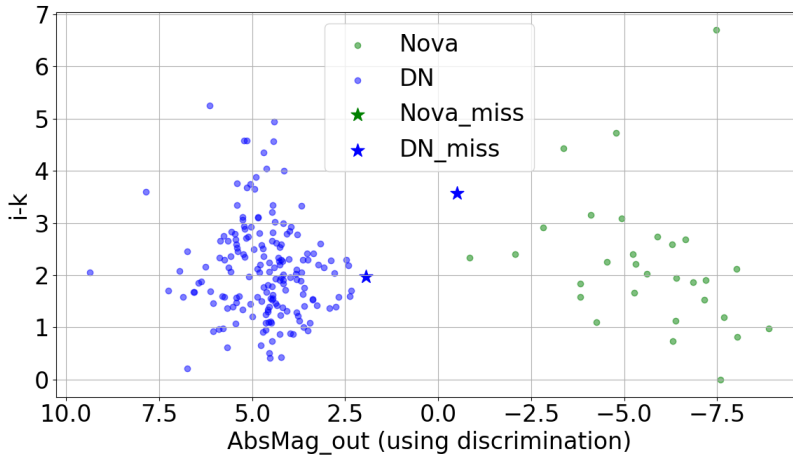


図 18: $AbsMag_{out}$ による判別結果を含めた $AbsMag_{out}, i-k$ の散布図

図17において、 $AbsMag_{out} = 2.5$ 付近の緑星印に注目する。これは、 $AbsMag_{out}$ のみでの判別の結果、本当は新星なのに矮新星であると誤判別したサンプルである。図18で同じサンプルを探すと、 $AbsMag_{out} = 2.5$ 付近の緑星印が存在しないことがわかる。これは、この緑星印のサンプルには $i-k$ の特徴量がないことを示す。つまり、 $AbsMag_{out}$, $Ampl$ の2つの特徴量を用いても誤判別してしまうサンプルに対して、近赤外線の情報である $i-k$ を加えることで正しく判別されるようになったわけではなく、誤判別したサンプルには近赤外線の情報がなく、単純に判別対象から消えてしまったということである。このように、近赤外線の情報を加えると $AUC=1.0$ を達成するのは、近赤外線の情報判別に有効であるというよりも、それまで誤判別されていたサンプルが消えるからだと考えられる。

また、誤判別するサンプルを調べる中で、教師データの誤りも見つかった。図17,18において、1つだけ $AbsMag_{out}$ が負の値になっている矮新星サンプルがある。このサンプルは KX Aql という矮新星である。SmartK では、教師データの距離は Gaia DR2 カタログを参照しており、ここでは KX Aql の年周視差は 0.2497 ± 0.2451 mas で、辛うじて検出できていた。これは距離でいうと $\frac{1}{0.2497} \sim 4.00$ kpc になる。ここで、距離と等級の関係について述べる。絶対等級 M は、天体を 10 pc の距離においた時の見かけの等級である。これを式で表すと、以下のようなになる。

$$M = m - 5 \log_{10}(d/10)$$

ここで、 m は見かけの等級、 d は地球から天体までの距離で、単位はパーセクである。1 パーセクは約 3.26 光年である。この式に、KX Aql の距離 4.00 kpc、見かけの等級 11.5 を代入すると、絶対等級 $M = -1.5$ となり、非常に明るくなる。ところが、2022 年、より精度の良い距離のデータ Gaia DR3 が出ており、それによると KX Aql の年周視差は 0.0957 ± 0.1948 mas になっている。これだと年周視差が有意に受かっているとは言えず、距離は不定になる。KX Aql は本来は普通の SU UMa 型矮新星なので [13]、距離は 100 pc ほどでそれほど遠くは無いはずである。それでも Gaia DR3 がこのような値なのは、天体が暗いためか、連星の公転運動が混ざっているからだと考えられる。この例のように、Gaia 衛星による距離データは DR3 によって精度が上がっているため、今後は DR3 で教師データを作り直す必要がある。

5.1 新星と矮新星

4.1 節の結果から、新星と矮新星の判別には、アウトバースト時の絶対等級 $AbsMag_{out}$ と、爆発振幅 $Ampl$ が有効であることが分かった。ただし、新星も矮新星も静穏時は同じような状態なので、 $AbsMag_{out}$ の違いが $Ampl$ に反映されていると考えられる。よって、爆発時の絶対等級の違いが、この2天体の判別に有効である。新星と矮新星を分ける特徴量として爆発時の絶対等級が選択された理由を以下で考察する。

新星爆発は白色矮星表面での核融合によって、降着していたガスが吹き飛ばされるため、その光度は概算としてエディントン光度で与えられる。エディントン光度とは、中心天体に周囲から物質が落ち込む球対称降着において、放射によ

る力と重力のつり合いで決まる限界光度のことで、以下で表される。

$$L_E = \frac{4\pi cGMm_p}{\sigma_T} = 1.2 \times 10^{31} \left(\frac{M}{M_\odot} \right) [\text{W}]$$

ここで、 $c, G, m_p, \sigma_T, M_\odot$ はそれぞれ光速、万有引力定数、陽子の質量、トムソン散乱の断面積、太陽質量である。激変星の白色矮星の典型的な質量は $0.6\text{--}1.38M_\odot$ なので、その時のエディントン光度は $L_E = 7.2\text{--}16.6 \times 10^{30} \text{ W}$ となる。これは、絶対等級へ変換すると約 $-6.55\text{--}5.65$ 等級となる。図 20 を見ても分かるように、新星の絶対等級は -10 から 2 等級なので、エディントン光度と矛盾しないことが分かる。対して、矮新星の絶対等級は 2 から 10 等級である。矮新星は降着円盤の増光であり、新星のように全体が吹き飛ばすほどの光度ではないため、新星よりも $AbsMag_{out}$ は暗くなる。これが、新星と矮新星の判別において $AbsMag_{out}$ ないしは $Ampl$ が有効である理由だと考えられる。

5.2 新星と WZ Sge 型星

4.2 節の結果から、新星と WZ Sge 型星の判別には、アウトバースト時の絶対等級 $AbsMag_{out}$ と、静穏時の絶対等級 $AbsMag_{qui}$ が有効であることが分かった。 $AbsMag_{out}$ についてはやはり、5.1 節と同様、新星は全体を吹き飛ばすほどの光度を見せるが、WZ Sge 型星のような矮新星は降着円盤の増光なので、そこで明るさに違いが出ると考えられる。 $AbsMag_{qui}$ については、伴星からの質量輸送率が関係しているのではないかと考えられる。図 19 に、降着円盤の平均絶対等級 (縦軸) と連星軌道周期 (横軸) を表す。軌道周期が長い激変星では、伴星からの放射と、伴星から流れ出たガスが降着円盤と衝突する地点であるホットスポットからの放射が卓越する。一方、軌道周期が短い激変星では、伴星は暗くなり、ホットスポットからの放射が卓越する。このホットスポットからの放射は、質量輸送率に比例する。さらに、激変星は連星系からの角運動量抜き取り機構によって、軌道周期が長い方から短い方、伴星が大きい系から小さい系へと進化し、質量輸送率は大きい系から小さい系へと進化する [14]。つまり、新星が軌道周期が長く質量輸送率が高い系であるのに対して、WZ Sge 型星は軌道周期が短く質量輸送率が低い系であるため、静穏時の絶対等級 $AbsMag_{qui}$ がこの 2 クラスの判別に有効なのだと考えられる。5.1 節の新星と矮新星において、静穏時の絶対等級 $AbsMag_{qui}$ が有効な特徴量として選ばれなかったのは、新星と矮新星の軌道周期や質量輸送率が同程度であるためだと考えられる。

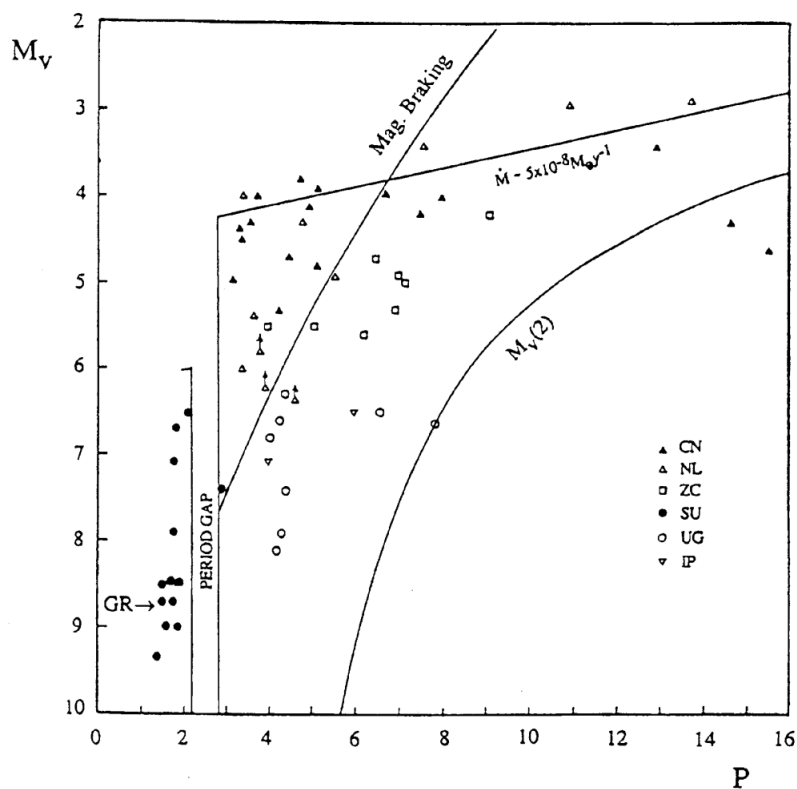


図 19: 降着円盤の平均絶対等級 M_V (縦軸) と連星軌道周期 (横軸) P を表す。伴星の絶対等級は、軌道周期の関数として $M_V(2)$ と表されている [14]。

5.3 新星とミラ型星

4.3 節の結果から、新星とミラ型星の判別には、爆発振幅 $Ampl$ 、静穏時の絶対等級 $AbsMag_{qui}$ 、可視光の色 $g-r, r-i, i-z$ 、近赤外における静穏時の絶対等級 $AbsMag_J$ が有効であることが分かった。 $Ampl, AbsMag_{qui}$ については、新星とミラ型星は極大光度が似ているため、静穏時の絶対等級や、そこからの差である爆発振幅が有効なのだと考えられる。 $g-r, r-i, i-z$ については、新星とミラ型星で可視光の色が違うため有効なのだと考えられる。例として、図 20 に HR 図 (横軸: $g-r$ 、縦軸: $AbsMag_{out}$) を示す。図 20 を見ると、 $g-r = 1.0$ 付近ではミラ型星を誤判別したサンプルがあるが、多くのサンプルで離れて分布していることが分かる。これは、 $r-i, i-z$ でも同様であった。

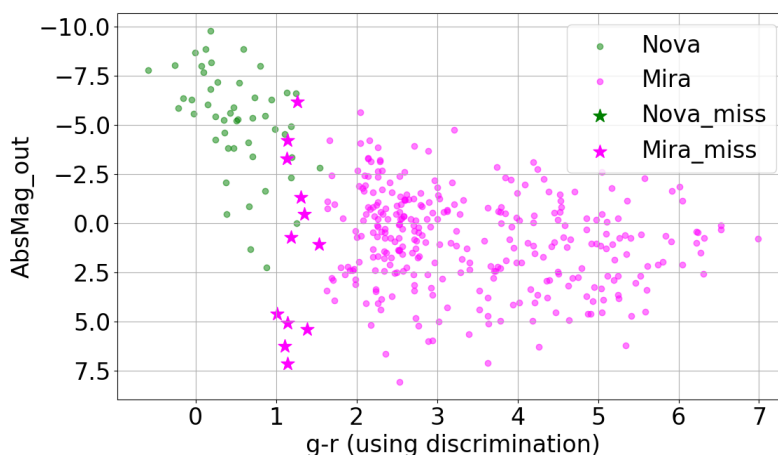


図 20: $g-r$ による判別結果を含めた $g-r$, $AbsMag_{out}$ の散布図 (HR 図)

また、今回の判別においては、近赤外線の情報以外の特徴量が追加された場合でも、 $AUC=1.0$ を達成したモデルがあった。表 9 のパターン 1 では天体までの距離 d_{kpc} を用いることで、完全分離可能なモデルを実現している。しかし、この特徴量は上述の近赤外線の特徴量と同様、サンプルを減らす効果によって完全分離を可能にしていることが確認できた。そのため、 d_{kpc} は新星とミラ型星に有効な特徴量とは言えない。このように、近赤外線の情報以外を追加して $AUC=1.0$ になった場合でも、追加する前に誤判別されたサンプルが判別対象から消えるということが起こる。

5.4 矮新星と WZ Sge 型星

4.4 節の結果から、矮新星と WZ Sge 型星の判別には、アウトバースト時の絶対等級 ($AbsMag_{out}$)、静穏時の絶対等級 ($AbsMag_{qui}$)、爆発振幅 ($Ampl$) のうち、少なくとも 2 つの特徴量が揃った時に、高い判別精度を出すことが分かった。また、爆発振幅 ($Ampl$) が利用可能な時は、距離関係 (d_{kpc} , gal_{abs}_z) のどちらか 1 つが加われば高い判別精度を出した。このことから、爆発振幅 $Ampl$ がこの 2 クラス判別において重要な役割を担っていると考えられる。 $Ampl$ で判別した場合の $AbsMag_{out}$ との散布図 (図 21) と、 $AbsMag_{qui}$ との散布図 (図 22) を示す。図 21, 22 から、 $Ampl$ においては WZ Sge 型星の方が大きいことが分かる。 $AbsMag_{out}$ においては WZ Sge 型星、 $AbsMag_{qui}$ においては矮新星が少し明るいことも分かり、 $Ampl$ のみだと誤判別してしまうサンプルを、この僅かな分布の違いによって正しく判別しているのではないかと考えられる。

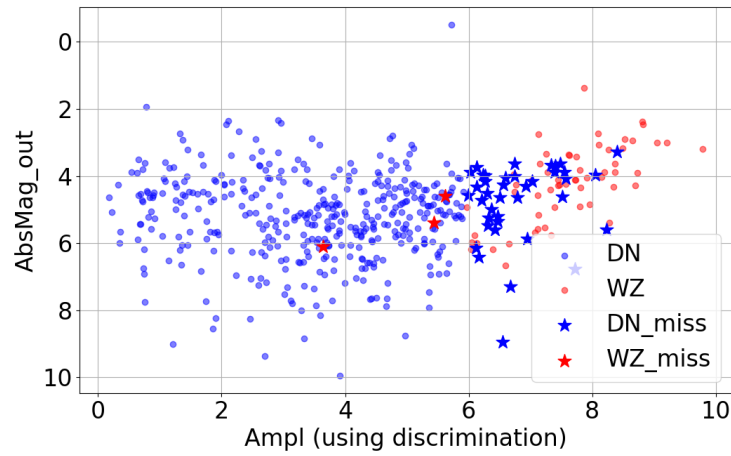


図 21: $Ampl$ による判別結果を含めた $Ampl, AbsMag_{out}$ の散布図

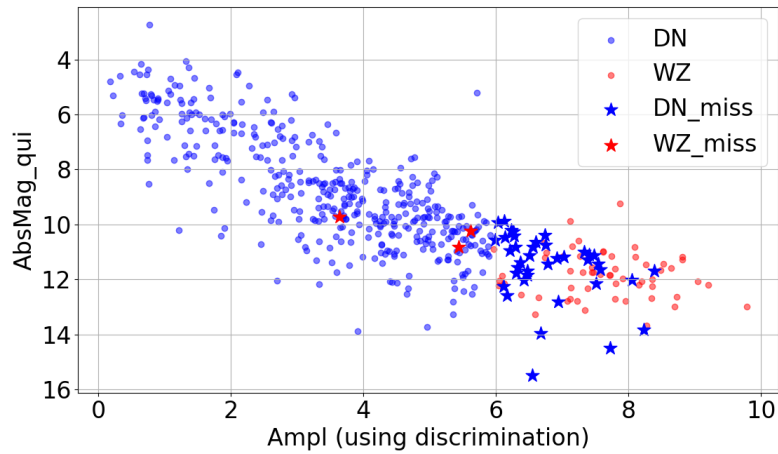


図 22: $Ampl$ による判別結果を含めた $Ampl, AbsMag_{qui}$ の散布図

WZ Sge 型星は矮新星のなかで最も軌道周期が短く、伴星からの質量輸送率が小さい系であるため、静穏時の絶対等級が矮新星のなかで最も暗いことは以前から知られていた。したがって、 $AbsMag_{qui}$ が選択されたことは自然に理解できる。一方、アウトバースト時の絶対等級は降着円盤からの黒体放射が卓越するため、両者の絶対等級は同程度と考えられてきた。本研究の結果は、アウトバースト時の絶対等級が WZ Sge 型星と通常の矮新星を判別する有効な特徴量であることを意味しており、WZ Sge 型星の方がアウトバースト時に降着円盤が高光度になることが明らかになった。これは、WZ Sge 型星のアウトバースト時には白色矮星への質量降着率が通常の矮新星よりも高いことを意味し、WZ Sge 型星に提

案されている 2:1 共鳴シナリオ [15] の証拠になる可能性がある。

2:1 共鳴とは、降着円盤中の物質が白色矮星まわりを 1 周する周期が、激変星の連星軌道周期の 1/2 になる位置で起こる共鳴現象のことで、その時の降着円盤の半径を 2:1 共鳴半径という。降着円盤の大きさが 2:1 共鳴半径に達した時、降着円盤が軸方向に膨らむことで、早期スーパーハンプのような振動現象が起こると提案されている。一方、通常の矮新星 (特に SU UMa 型) では 3:1 共鳴が起こる。降着円盤が 3:1 共鳴半径に達すると、降着円盤が楕円形に変化して歳差運動を起こし、早期スーパーハンプと似た振動現象である通常のスーパーハンプを起こす。矮新星のアウトバースト時に降着円盤が到達できる最大半径は連星質量比によって決まる。矮新星のなかで SU UMa 型は、連星質量比が $q = M_2/M_1 < 0.25$ で 3:1 共鳴半径に達することができ、さらに q が小さい WZ Sge 型は 2:1 共鳴半径に達することができると考えられている。白色矮星への質量降着率は角運動量の外側への輸送率で決まるのだが、WZ Sge 型星のアウトバースト時の絶対等級が、通常の矮新星と比べて大きいということは、WZ Sge 型星には通常の矮新星にはない角運動量輸送機構があるはずである。その候補は现阶段では 2:1 共鳴しかないため、今回の結果は 2:1 共鳴シナリオを支持する。この仮説の精査は将来の研究課題である。

5.5 矮新星とフレア星

4.5 節の結果から、矮新星とフレア星の判別において、可視光の色 ($g-r$) が有効であることが分かった。図 23 に HR 図 (横軸: $g-r$ 、縦軸: $AbsMag_{out}$) を示す。図 23 を見ると、黄色点で表されたフレア星の分布に 2 つの集団が見られる。 $g-r = 1.0-1.5$ 、 $AbsMag_{out} = 10.0$ 前後の集団は、K5 型から M 型の主系列星である。一方、 $g-r = 0.5$ 、 $AbsMag_{out} = 5.0$ 前後の集団は、G 型の主系列星である。ここで、K5 型や M 型、G 型というのは、スペクトル分類によって分類された恒星のスペクトル型のこと、温度、つまり色 $g-r$ に対応している。G 型の方が、K5 型や M 型と比べて温度が高く青色を示す。さらに、矮新星の伴星とフレア星は同じ恒星である。しかし、矮新星にはホットスポットや降着円盤などの伴星よりも温度が高い、つまり青い放射成分がある。そのため、同じような恒星を含んでいても全体としてフレア星とは異なる色を示し、 $g-r$ が判別に有効な特徴量になるのだと考えられる。一方で、矮新星とフレア星で増光機構は全く異なるので、両者の $AbsMag_{out}$ は異なる分布になると期待されるが、色が青い、図 23 における $g-r = 0.5$ まわりのフレア星、つまり G 型の主系列星の $AbsMag_{out}$ は矮新星と同等の値をもつことが分かる。G 型の主系列星の半径が約 10^{11} cm で、矮新星の降着円盤の大きさもこれと同程度なので、両者とも似た温度で黒体放射で光るなら、光度も似た分布になる。これによって、矮新星とフレア星の判別に $AbsMag_{out}$ が有効にはならないのだと考えられる。

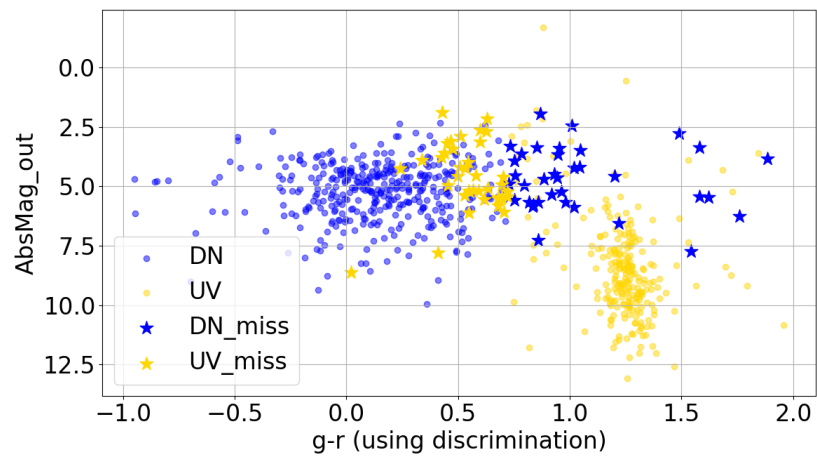


図 23: $g-r$ による判別結果を含めた $g-r, AbsMag_out$ の散布図 (HR 図)

6 まとめ

本論文では、新星・矮新星などの激変星を研究対象とした。激変星は突発的に増光するため、その極初期に適切な観測を行うことが難しく、未解明な物理が多く潜んでいる。

そこで、そのような突発現象に対して、天体の型判別や適切な追跡観測を自動で行うシステム SmartK を開発することで、激変星についての理解を進めることを目指した。

SmartK では、天体の型を判別するために機械学習を用いた。本論文ではこの機械学習についての調査が主題である。今回用いた教師データには欠損値が多いため、教師データの扱い方が異なる 3 つの判別モデルを用いることで、SmartK に適切な判別モデルはどのモデルなのかを評価した。そして、本論文で用いた教師データにおいては非線形な境界は必要なく、正規分布を仮定した生成モデルが適切であるという結果を得た。

本論文全体として、判別に有効な特徴量をデータから明らかにすることで、判別モデルの高精度化と共に天体の理解を深める、という目的があった。生成モデルを用いて明らかになった、判別に有効な特徴量を表 11 にまとめる。

判別天体	有効な特徴量
新星と矮新星	$AbsMag_{out}, Ampl$
新星と WZ Sge 型星	$AbsMag_{out}, AbsMag_{qui}$
新星とミラ型星	$Ampl, AbsMag_{qui}, g - r, r - i, i - z$
矮新星と WZ Sge 型星	$AbsMag_{out}, AbsMag_{qui}, Ampl$
矮新星とフレア星	$g - r$

表 11: 判別に有効な特徴量

ここに示された計 6 個の特徴量が利用可能であれば、全ての特徴量が揃っていなくても、 $AUC > 0.95$ の精度で判別できることが分かった。過適合を防ぐという意味で、SmartK においても利用する特徴量を限定していくことなどは、これからの課題である。

また、天体への理解も深まった。矮新星と WZ Sge 型星の判別に $AbsMag_{out}$ が有効であると分かったことで、これまで両者のアウトバースト時の絶対等級は同程度であると考えられていたがそうとは限らず、WZ Sge 型星の方がアウトバースト時には白色矮星への質量降着率が通常の矮新星よりも高い可能性が示唆された。

References

- [1] 中村泰久、2009、現代天文学第7巻『恒星』、日本評論社
- [2] V339 Del Photometry <http://www.astro.sunysb.edu/fwalter/SMARTS/NovaAtlas/v339del/phot/v339del>.
- [3] 田口 他、2020、せいめいユーザーズミーティング、
<http://seimei.nao.ac.jp/files/UM/2020/Oral/Taguchi.pdf>
- [4] 田口、2022、激変星研究会、<http://www.kusastro.kyoto-u.ac.jp/iso-gai/cvs2022/slides/taguchi.pdf>
- [5] 植村誠、2017、岩波データサイエンス Vol.6、岩波書店
- [6] Uemura et al.,2021,PASJ,vol.64,p.92
- [7] AAVSO,Mira Variables with Period Changes,<https://www.aavso.org/mira-variables-period-changes>
- [8] Bopp & Moffett,1973,The Astrophysical Journal,Vol.185,p.243
- [9] 古賀、2021、卒業論文
- [10] 広島大学宇宙科学センター、概要、https://www.hiroshimau.ac.jp/hasc/institution_jp/telescope/abstract
- [11] ひろ、ベイズの定理とは？誰でも理解できるようにわかりやすく解説、
<https://www.headboost.jp/bayes-theorem/>
- [12] 植村誠、2023(出版予定)
- [13] Kato et al.,2010,PASJ,3.1 節
- [14] Warner, B. 1987, MNRAS, 227, 23
- [15] Osaki,Y.& Meyer,F. 2002, A&A 383,574

A 付録

A.1 混同行列

3.1 節で計算した、判別モデルごと、特徴量の組み合わせごとの混同行列を表 12-35 に示す。

A.1.1 LR

正解率:28.77%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	176	4	3	111	4
矮新星	462	86	215	175	103
WZ Sge 型星	54	17	38	33	23
ミラ型星	336	23	66	309	45
フレア星	81	27	83	61	169

表 12: 座標 (2)

正解率:64.18%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	93	1	4	14	1
矮新星	0	340	191	24	120
WZ Sge 型星	0	12	61	2	1
ミラ型星	84	52	36	378	6
フレア星	2	59	8	8	248

表 13: 座標、距離 (5)

正解率:91.66%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	88	1	8	0	0
矮新星	1	504	63	0	16
WZ Sge 型星	5	7	68	0	0
ミラ型星	0	0	0	349	5
フレア星	2	8	0	3	299

表 14: 座標、静穏時可視 (6)

正解率:65.93%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	17	31	18	14	7
矮新星	48	158	50	4	42
WZ Sge 型星	4	2	0	0	4
ミラ型星	39	0	9	723	7
フレア星	106	34	110	1	128

表 15: 座標、静穏時近赤外 (4)

正解率:94.16%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	48	0	0	0	0
矮新星	0	358	42	0	2
WZ Sge 型星	0	5	53	0	0
ミラ型星	0	0	0	257	2
フレア星	2	8	0	0	268

表 16: 座標、距離、静穏時可視 (10)

正解率:90.93%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	56	1	1	1	0
矮新星	2	216	46	0	18
WZ Sge 型星	0	5	4	0	0
ミラ型星	3	0	0	552	1
フレア星	2	29	2	0	286

表 17: 座標、距離、静穏時近赤外 (8)

正解率:91.56%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	28	0	4	0	0
矮新星	5	127	27	0	9
WZ Sge 型星	2	1	3	0	0
ミラ型星	0	0	0	351	3
フレア星	2	15	5	0	283

表 18: 座標、静穏時可視、静穏時近赤外 (9)

正解率:94.06%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	28	0	0	0	0
矮新星	0	132	19	0	6
WZ Sge 型星	0	3	3	0	0
ミラ型星	0	0	0	258	1
フレア星	1	10	3	0	260

表 19: 全部 (14)

A.1.2 SMLR

正解率:36.53%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	202	5	6	74	11
矮新星	154	137	227	314	209
WZ Sge 型星	15	24	47	27	52
ミラ型星	256	39	35	387	62
フレア星	23	24	89	70	215

表 20: 座標 (2)

正解率:60.63%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	95	0	3	14	1
矮新星	0	198	199	42	236
WZ Sge 型星	0	4	59	3	10
ミラ型星	63	21	33	419	20
フレア星	1	18	8	11	287

表 21: 座標、距離 (5)

正解率:89.83%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	88	1	8	0	0
矮新星	2	485	80	0	17
WZ Sge 型星	4	9	67	0	0
ミラ型星	0	0	0	346	8
フレア星	0	14	0	2	296

表 22: 座標、静穏時可視 (6)

正解率:66.83%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	45	17	11	7	7
矮新星	63	119	51	4	65
WZ Sge 型星	1	3	2	0	4
ミラ型星	79	3	20	671	5
フレア星	32	46	93	5	203

表 23: 座標、静穏時近赤外 (4)

正解率:93.30%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	46	1	1	0	0
矮新星	2	349	43	0	8
WZ Sge 型星	0	5	53	0	0
ミラ型星	2	0	0	256	1
フレア星	1	6	0	0	271

表 24: 座標、距離、静穏時可視 (10)

正解率:88.16%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	55	2	2	0	0
矮新星	5	198	58	1	20
WZ Sge 型星	0	3	4	0	2
ミラ型星	5	0	4	547	0
フレア星	2	28	11	2	276

表 25: 座標、距離、静穏時近赤外 (8)

正解率:93.52%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	29	1	2	0	0
矮新星	8	145	12	0	3
WZ Sge 型星	2	3	1	0	0
ミラ型星	0	0	0	352	2
フレア星	0	15	7	1	282

表 26: 座標、静穏時可視、静穏時近赤外 (9)

正解率:95.58%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	26	1	0	1	0
矮新星	0	143	11	0	3
WZ Sge 型星	0	5	1	0	0
ミラ型星	2	0	0	256	1
フレア星	0	7	1	0	266

表 27: 全部 (14)

A.1.3 GM

正解率:28.66%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	266	7	7	13	5
矮新星	340	165	141	165	230
WZ Sge 型星	22	33	34	24	52
ミラ型星	527	40	34	130	48
フレア星	67	33	63	78	180

表 28: 座標 (2)

正解率:61.03%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	94	0	3	15	1
矮新星	0	302	249	26	98
WZ Sge 型星	0	14	58	3	1
ミラ型星	89	29	81	354	3
フレア星	2	36	15	15	257

表 29: 座標、距離 (5)

正解率:88.92%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	91	1	5	0	0
矮新星	4	466	104	0	10
WZ Sge 型星	3	9	68	0	0
ミラ型星	0	0	0	348	6
フレア星	2	13	0	1	296

表 30: 座標、静穏時可視 (6)

正解率:77.76%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	49	16	5	8	9
矮新星	43	145	42	1	71
WZ Sge 型星	2	2	0	1	5
ミラ型星	60	0	18	697	3
フレア星	10	30	17	3	319

表 31: 座標、静穏時近赤外 (4)

正解率:88.42%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	45	2	0	1	0
矮新星	3	312	83	0	4
WZ Sge 型星	0	5	53	0	0
ミラ型星	1	0	0	257	1
フレア星	3	18	0	0	257

表 32: 座標、距離、静穏時可視 (10)

正解率:81.55%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	53	1	2	3	0
矮新星	2	123	141	1	15
WZ Sge 型星	1	1	7	0	0
ミラ型星	2	0	0	554	0
フレア星	2	32	21	2	262

表 33: 座標、距離、静穏時近赤外 (8)

正解率:94.56%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	21	3	8	0	0
矮新星	0	146	15	0	7
WZ Sge 型星	1	1	3	0	1
ミラ型星	0	0	0	351	3
フレア星	0	8	0	0	297

表 34: 座標、静穏時可視、静穏時近赤外 (9)

正解率:91.29%	新星	矮新星	WZ Sge 型星	ミラ型星	フレア星
新星	28	0	0	0	0
矮新星	0	113	41	0	3
WZ Sge 型星	0	2	3	0	1
ミラ型星	0	0	0	259	0
フレア星	0	16	0	0	258

表 35: 全部 (14)

A.2 正解率

3.2 節で計算した、判別モデルごと、特徴量の組み合わせごとの正解率を図 24-31 に示す。

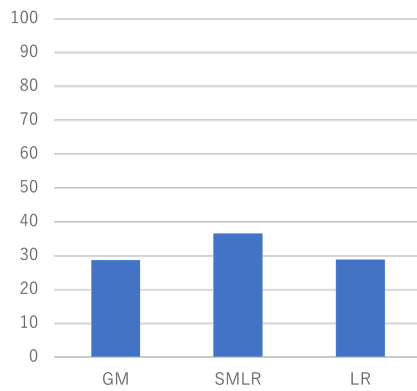


図 24: 座標 (2)

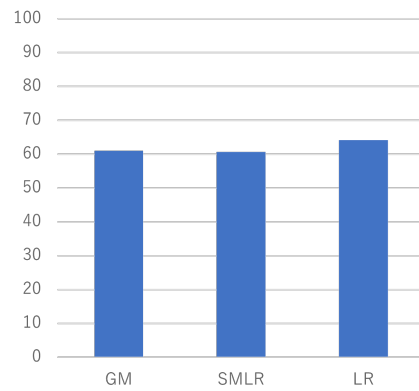


図 25: 座標、距離 (5)

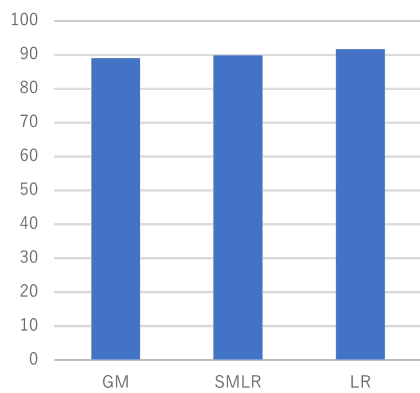


图 26: 座標、静穩時可視 (6)

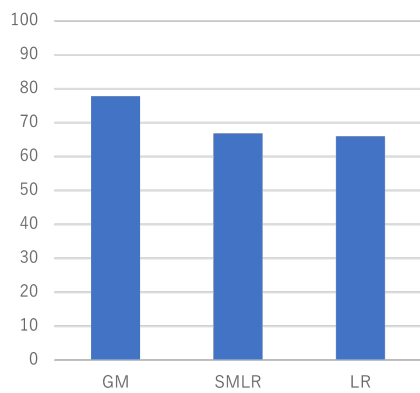


图 27: 座標、静穩時近赤外 (4)

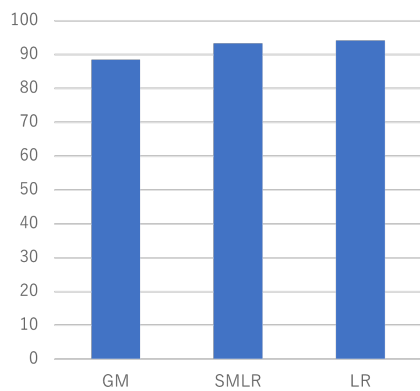


图 28: 座標、距離、静穩時可視 (10)

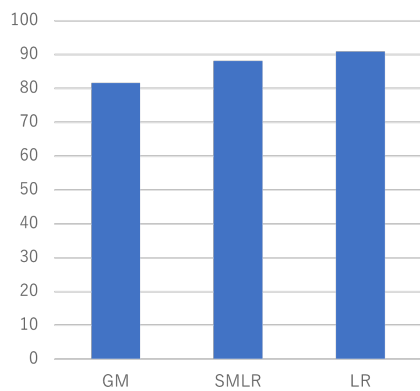


图 29: 座標、距離、静穩時近赤外 (8)

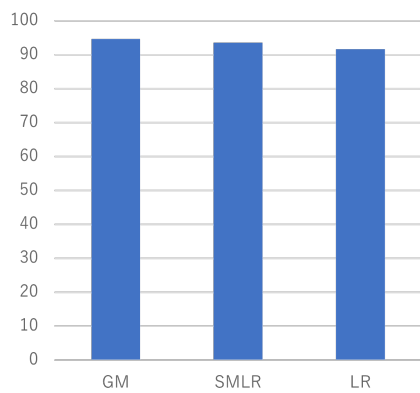


図 30: 座標、静穏時可視、静穏時近赤外 (9)

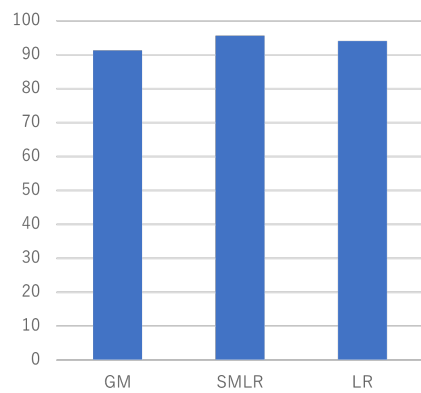


図 31: 全部 (14)

謝辞

まず、指導教員の植村誠准教授。学部生の頃から約2年半、お忙しいなか大変分かりやすい指導や質問への丁寧な回答など、本当にお世話になりました。学会や研究会への参加も、貴重な経験になりました。改めて、本当にありがとうございました。また、研究室の先輩方、同級生、後輩も、メンバーに恵まれとても楽しい研究室生活でした。研究室メンバーの大半で行った新潟での学会は、もはや新潟旅行でした。同級生5人で行った四国旅行もいい思い出ですねえ。就職してもまた皆さんと会いたいです。

そして、中村謙吾くん。M1の夏に行った東京のインターンが偶々同じで、そこからここまで急接近(?)するとは思いませんでした。さらに就職先も近い場所という奇跡。社会人になってもよろしく願いいたします。

謝辞はある程度ふざけても大丈夫な場所だと聞いていたのもっと色々書きたかったですが、あまり思いつかないのでこの辺りで終わろうかと思います。