

W42b 欠損の多い教師データを用いた銀河系内突発現象の機械判別

古賀柚希, 植村誠, 佐崎凌佑 (広島大学), 池田思朗 (統計数理研究所)

新星爆発や矮新星アウトバーストなどの突発的に起こる現象を効率的に研究するには、迅速に適切な追跡観測を自動で行う必要があります、そのためには、似た振る舞いをする他の天体と目的天体を区別しなければならない。これには機械学習が有用である。我々は、新星、矮新星、WZ Sge型矮新星、ミラ型変光星、フレア星の5つの型を想定し、座標と距離、静穏時の可視光と近赤外線の等級から得られる14の特徴量を教師データとして、機械判別を行ってきた。しかし、この教師データには欠損値が多く、現在使っている「スパース多クラスロジスティック回帰 (SMLR)」という判別モデルは、非線形で複雑な決定境界を定めることができるが、全特徴量が揃っているサンプルしか扱えない。このような欠損の多い教師データを使わざるを得ないことは、天文学の他のケースでも生じる共通の問題である。

そこで、我々はベイズの定理の表式をそのまま使う「生成モデル (GM)」を扱い、様々な特徴量の組み合わせに対して、SMLR と判別性能を比較した。今回は尤度関数を多変量正規分布とし、教師データの分散・共分散を用いた。欠損の多いデータなので、分散共分散行列が半正定値行列にならないことがあるが、分散の推定値に簡単な補正を加えることで、全サンプルを教師データとして用いることができるようになった。SMLR と GM について、交差検証で得られた正解率に大きな差はなく、今回のデータに対しては複雑な決定境界が不要であることが示唆された。また、静穏時の近赤外線のデータをもつ WZ Sge 型矮新星のサンプルが少ないため、同程度の性能であれば、より多くのデータを使える GM の方が安定した結果を期待できることや、計算時間の観点から常に GM を使った方がよいこと等も結果として得られた。