

Classificação de objetos astronômicos de brilho variável utilizando suas curvas de luz

Luigi Lucas de Carvalho Silva

Centro Brasileiro de Pesquisas Físicas – CBPF
Ministério da Ciência, Tecnologia e Inovações – MCTI
Rio de Janeiro - RJ - Brasil

luigi.lcsilva@gmail.com

13 de dezembro de 2021

- 1 Introdução
- 2 Situação-Problema
- 3 Objetivos
- 4 Fundamentação Teórica
 - Random Forest
 - Métricas
- 5 Tratamento dos Dados e Aplicação do Modelo
- 6 Resultados
 - Resultados

- 1 Introdução
- 2 Situação-Problema
- 3 Objetivos
- 4 Fundamentação Teórica
- 5 Tratamento dos Dados e Aplicação do Modelo
- 6 Resultados

- Verifica-se a existência de diversos objetos astronômicos de brilho variável. Exemplos: estrelas variáveis, supernovas, estrelas binárias, trânsito de exoplanetas, dentre outros.
- Curvas de brilho por tempo desses objetos: curvas de luz.
- Cada objeto terá uma curva de luz com características próprias.

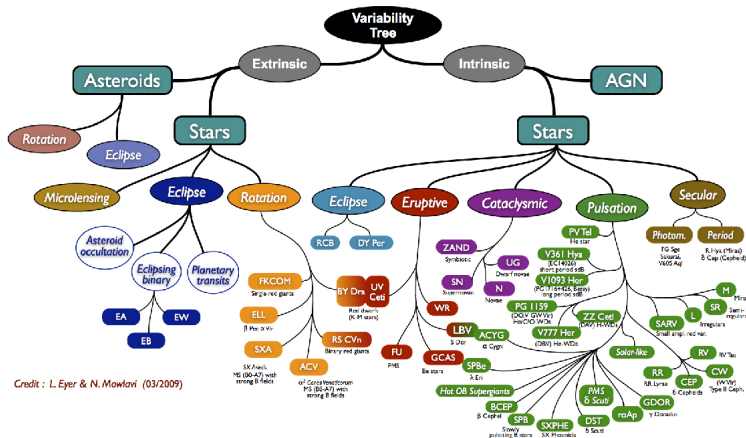


Figura: Diferentes tipos de objetos astronômicos de brilho variável [2].

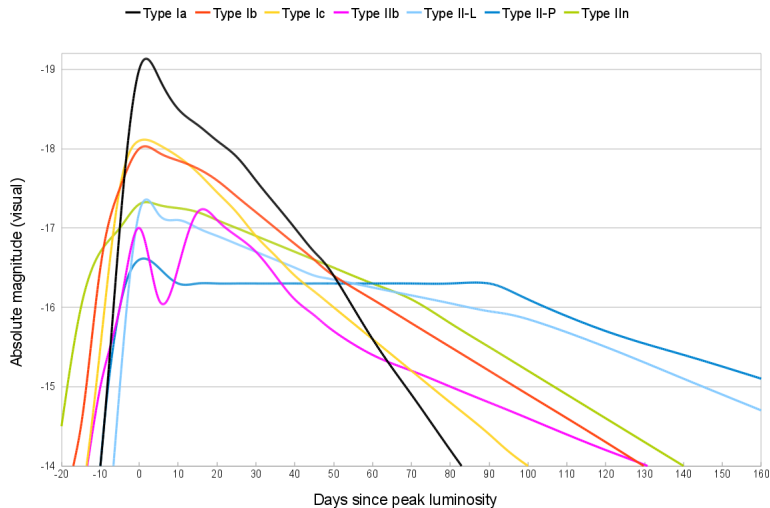


Figura: Exemplos de curvas de luz de supernovas [3].

- 1 Introdução
- 2 Situação-Problema**
- 3 Objetivos
- 4 Fundamentação Teórica
- 5 Tratamento dos Dados e Aplicação do Modelo
- 6 Resultados

- Large Synoptic Survey Telescope (LSST): expectativa de identificar muitos objetos de brilho variável desses tipos (uma ordem de grandeza ou mais) [2].
- Desafio de como classificar esses objetos, classificação essa necessária para diferentes estudos científicos.
- “Os dados do LSST provavelmente responderão perguntas que ainda nem sabemos fazer.” [2]

- Hoje, um dos métodos mais precisos para classificar esses objetos é a espectroscopia.
- Problemas: tempo de telescópio, quantidade de objetos.
- Desafio Photometric LSST Astronomical Time-Series Classification Challenge (PLAsTiCC); pode ser acessado na plataforma Kaggle [4].

- 1 Introdução
- 2 Situação-Problema
- 3 Objetivos**
- 4 Fundamentação Teórica
- 5 Tratamento dos Dados e Aplicação do Modelo
- 6 Resultados

- Mostrar algumas ideias para a resolução desse problema de classificação.
- Modelo utilizado: Random Forest.
- Os dados utilizados foram obtidos na plataforma do desafio [2].

- 1 Introdução
- 2 Situação-Problema
- 3 Objetivos
- 4 Fundamentação Teórica**
 - Random Forest
 - Métricas
- 5 Tratamento dos Dados e Aplicação do Modelo
- 6 Resultados

- Modelo mais simples, que precede o *Random Forest: Decision Trees*.
- Dadas certas características, ou *features*, pode-se filtrar os dados, ramo a ramo, até chegar-se em uma classificação para os objetos.

Decision Tree

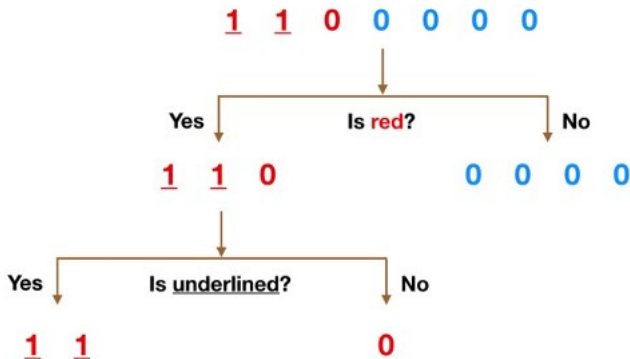


Figura: Exemplo simples de árvore de decisão [7].

- O modelo *Random Forest* nada mais é do que várias *Decision Trees*.
- Cada uma faz sua própria predição e, ao final, a classe mais votada é dada como nossa predição final.

- 1 Introdução
- 2 Situação-Problema
- 3 Objetivos
- 4 Fundamentação Teórica**
 - Random Forest
 - Métricas
- 5 Tratamento dos Dados e Aplicação do Modelo
- 6 Resultados

- Existem várias métricas para medir o quão bom é o modelo [5]. Aqui, serão exibidos dois exemplos:
- Classification Accuracy
- Precision

- 1 Introdução
- 2 Situação-Problema
- 3 Objetivos
- 4 Fundamentação Teórica
- 5 Tratamento dos Dados e Aplicação do Modelo
- 6 Resultados

- A discussão será feita acompanhando-se o notebook.

- 1 Introdução
- 2 Situação-Problema
- 3 Objetivos
- 4 Fundamentação Teórica
- 5 Tratamento dos Dados e Aplicação do Modelo
- 6 Resultados
 - Resultados

- Com o método de extração de *features* utilizado e o modelo escolhido, a acurácia ficou em 67%.
- Possíveis caminhos para a solução:
- Tratar os dados de maneiras distintas. Por exemplo, de maneira periódica e ordenada, ao invés de geral. Tratamento usado em exemplos do sktime [6];
- Usar a razão sinal ruído, ou outro método que leve em consideração os erros no fluxo medido;
- Utilização de outros modelos [1].



AMIDON, A.

A Brief Survey of Time Series Classification Algorithms.

<https://towardsdatascience.com/a-brief-introduction-to-time-series-classification-algorithms>
2020.

[Acessado em 07/12/2021].



APERS, M.

The PLAsTiCC Astronomy “Starter Kit”.

<https://www.kaggle.com/michaelapers/the-plasticc-astronomy-starter-kit/notebook>, 2018.

[Acessado em 05/12/2021].



LITHOPSIAN.

File:Comparative supernova type light curves.png.

https://en.m.wikipedia.org/wiki/File:Comparative_supernova_type_light_curves.png, 2012.

[Acessado em 05/12/2021].



LSST-PROJECT.

PLAsTiCC Astronomical Classification.

<https://www.kaggle.com/c/PLAsTiCC-2018>, 2018.

[Acessado em 05/12/2021].



MINAEE, S.

20 Popular Machine Learning Metrics. Part 1: Classification and Regression Evaluation Metrics.

<https://towardsdatascience.com/>

20-popular-machine-learning-metrics-part-1-classification-2019.

[Acessado em 07/12/2021].



SKTIME-DOCUMENTATION.

Univariate time series classification with sktime.

https://www.sktime.org/en/stable/examples/02_classification_univariate.html, 2021.

[Acessado em 07/12/2021].



YIU, T.

Understanding Random Forest.

[https://towardsdatascience.com/](https://towardsdatascience.com/understanding-random-forest-58381e0602d2)

[understanding-random-forest-58381e0602d2](https://towardsdatascience.com/understanding-random-forest-58381e0602d2), 2019.

[Acessado em 05/12/2021].