

CLASSIFICAÇÃO MORFOLÓGICA DE GALÁXIAS DO S-PLUS USANDO CONJUNTO DE REDES NEURAI CONVOLUCIONAIS

Natanael Magalhães Cardoso

Profa. Cláudia Mendes de Oliveira

Instituto de Astronomia, Geofísica e Ciências Atmosféricas, USP – SP

nauxmac@gmail.com

Objetivos

A classificação morfológica é a catalogação de galáxias de acordo com a sua aparência visual e a classificação está ligada com as propriedades físicas da galáxia. Uma classificação morfológica feita através de inspeção visual está sujeita a um viés causado pela subjetividade da observação humana. Por isso, a classificação sistemática, objetiva e facilmente reproduzível de galáxias vem ganhando importância desde quando o astrônomo Edwin Hubble criou seu famoso método de classificação [1]. Neste trabalho, nós combinamos classificações visuais acuradas do projeto GalaxyZoo¹ com métodos de *Deep Learning*. O objetivo é encontrar uma técnica automática e eficiente que consiga simular a classificação visual humana a partir da análise de imagens do DR1 do Southern Photometric Local Universe Survey (S-PLUS) [2] e publicar um catálogo com galáxias ainda não classificadas. Para isto, um modelo *Deep Learning* foi desenvolvido através de um conjunto de quatro redes neurais convolucionais usando a técnica *Stacking Ensemble* [3].

Métodos e Procedimentos

O modelo proposto classifica a morfologia das galáxias a partir de suas imagens. As imagens de cada uma das 12 bandas (detalhes sobre cada banda na Tabela 1 de [2]) foram obtidas a partir do banco de dados do projeto S-PLUS²,

agrupadas em 3 bandas e convertidas para o espaço de cor RGB. Esta etapa de redução permite usar informações dos 12 filtros do telescópio em imagens com apenas 3 bandas, causando redução no tempo de treino dos modelos. Como a técnica proposta consiste em um treinamento supervisionado, além dos dados a serem classificados, também são necessárias as classificações por humanos do GalaxyZoo para que o algoritmo consiga relacionar os padrões entre as imagens e suas respectivas classes. Foram criados os conjuntos de treinamento, validação e teste, com aproximadamente 30% de galáxias elípticas e 70% de galáxias espirais. Técnicas como reamostragem aleatória e ponderamento das classes foram usadas para tratar o desbalanceamento. Um outro conjunto, denominado *blind*, foi criado com galáxias ainda não classificadas pelo GalaxyZoo e que serão classificadas pelo modelo. Todos estes conjuntos possuem distribuição de *redshift* e magnitude similares.

Para treinar os classificadores, foram usadas redes neurais convolucionais com os pesos inicializados a partir de um pré-treinamento na ImageNet³. O pré-processamento dos dados, o ajuste dos hiper-parâmetros e o treinamento foi feito para cada classificador independentemente. Dos modelos treinados, os que obtiveram melhor performance de classificação são baseados nas arquiteturas InceptionResNetV2, EfficientNetB2, DenseNet-169 e DenseNet-201. Para aproveitar o potencial destes quatro classificadores, um outro modelo (denominado meta-modelo) foi treinado para receber as predições

¹<http://galaxyzoo.org>

²<http://splus.cloud>

³<http://image-net.org>

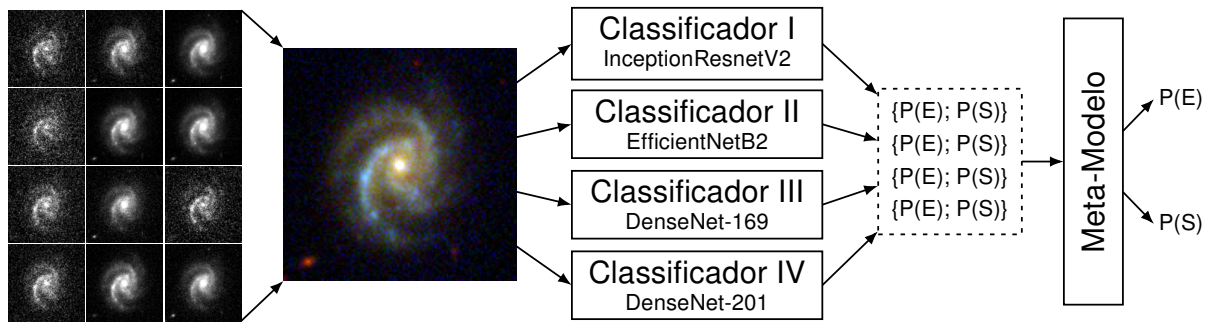


Figura 1: Diagrama da arquitetura. Da esquerda para a direita, as 12 imagens de cada banda são agrupadas em uma única imagem RGB, que é a entrada dos classificadores individuais. Estes classificadores têm a função de extrair características visuais da imagem e retornar a probabilidade de ser elíptica ou espiral. O meta-modelo tem a função de combinar as predições dos classificadores em uma única predição final mais robusta.

de cada classificador e retornar uma única predição. Formando, assim, um conjunto (*Ensemble*) de classificadores. Este procedimento é resumido no diagrama da Figura 1.

Resultados

Uma forma de avaliar o modelo quantitativamente é a partir da Curva Característica de Operação do Receptor (ROC) no conjunto de teste, como mostra a Figura 2. A partir deste gráfico é possível tirar conclusões sobre o potencial de classificação do modelo. A área abaixo da

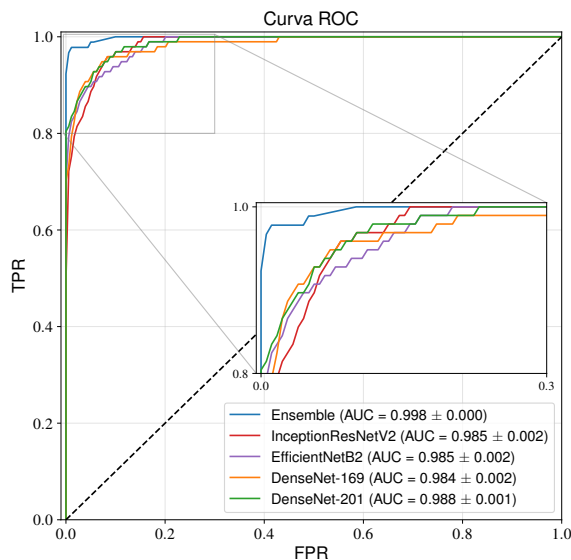


Figura 2: Curva ROC para cada classificador e para o *Ensemble*. Cada curva representa a mediana de 60 medições, a área abaixo da curva (AUC) e seu desvio padrão estão na legenda de cada curva.

curva, cujo valor máximo é 1, indica a probabilidade de classificação correta pelo modelo. Além disso, quanto mais rápida a curva chega ao topo, maior é o grau de separabilidade do modelo, ou seja, maior a capacidade do modelo de distinguir entre as classes. Como esperado, o *Ensemble* (curva azul), que é a união dos quatro classificadores, supera a performance dos classificadores individualmente, atingindo uma AUC de 0.998.

Conclusões

A partir deste projeto foram obtidos um modelo de *Deep Learning* completamente desenvolvido e operante com alto potencial de classificação morfológica entre galáxias elípticas e espirais a partir de imagens do S-PLUS, além de um catálogo com classificações automatizadas de 2536 galáxias com magnitude menor que 17 sem classificações no GalaxyZoo, ambos publicamente disponíveis⁴. Além disso, um artigo científico sobre este projeto encontra-se em etapa final de escrita.

Referências Bibliográficas

- [1] E. P. Hubble. "Extragalactic nebulae". Em: (dez. de 1926).
- [2] C. M. de Oliveira et al. "The Southern Photometric Local Universe Survey (S-PLUS): improved SEDs, morphologies, and redshifts with 12 optical filters". Em: (ago. de 2019).
- [3] D. H. Wolpert. "Stacked generalization". Em: (1992).

⁴<http://natanael.net>