



Universidade Federal  
do Rio de Janeiro  

---

Escola Politécnica

## IDENTIFICAÇÃO FOTOMÉTRICA DE SUPERNOVAS ATRAVÉS DE ALGORITMOS DE MACHINE LEARNING

Felipe Matheus Fernandes de Oliveira

Projeto de Graduação apresentado ao Curso de Engenharia de Controle e Automação da Escola Politécnica, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Engenheiro.

Orientadores: Amit Bhaya

Ribamar Rondon de Rezende  
dos Reis

Rio de Janeiro  
Julho de 2019

IDENTIFICAÇÃO FOTOMÉTRICA DE SUPERNovas ATRAVÉS DE  
ALGORITMOS DE MACHINE LEARNING

Felipe Matheus Fernandes de Oliveira

PROJETO DE GRADUAÇÃO SUBMETIDO AO CORPO DOCENTE DO  
CURSO DE ENGENHARIA DE CONTROLE E AUTOMAÇÃO DA ESCOLA  
POLITÉCNICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO  
PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU  
DE ENGENHEIRO DE AUTOMAÇÃO.

Examinado por:

---

Prof. Amit Bhaya, D.Sc.

---

Prof. Ribamar Rondon de Rezende dos Reis, Ph.D.

---

Prof. Heraldo Luís Silveira de Almeida, D.Sc.

---

Prof. Nome do Segundo Examinador Sobrenome, D.Sc.

RIO DE JANEIRO, RJ – BRASIL  
JULHO DE 2019

Fernandes de Oliveira, Felipe Matheus

IDENTIFICAÇÃO FOTOMÉTRICA DE  
SUPERNOVAS ATRAVÉS DE ALGORITMOS DE  
MACHINE LEARNING/Felipe Matheus Fernandes de  
Oliveira. – Rio de Janeiro: UFRJ/ Escola Politécnica,  
2019.

XIII, 19 p.: il.; 29, 7cm.

Orientadores: Amit Bhaya

Ribamar Rondon de Rezende dos Reis

Projeto de Graduação – UFRJ/ Escola Politécnica/  
Curso de Engenharia de Controle e Automação, 2019.

Referências Bibliográficas: p. 18 – 18.

1. Machine Learning. 2. Gaussian Process Fitting.
3. Supernova Photometric Identification. I. Bhaya,  
Amit *et al.* II. Universidade Federal do Rio de Janeiro,  
Escola Politécnica, Curso de Engenharia de Controle e  
Automação. III. Título.

*A alguém cujo valor é digno  
desta dedicatória. A você Goku*

# Agradecimentos

Gostaria de agradecer a todos os Sayajins por terem nos salvado das garras maléficas de Freeza, Cell e Majin Boo.

Resumo do Projeto de Graduação apresentado à Escola Politécnica/ UFRJ como parte dos requisitos necessários para a obtenção do grau de Engenheiro de Automação.

## IDENTIFICAÇÃO FOTOMÉTRICA DE SUPERNOVAS ATRAVÉS DE ALGORITMOS DE MACHINE LEARNING

Felipe Matheus Fernandes de Oliveira

Julho/2019

Orientadores: Amit Bhaya

Ribamar Rondon de Rezende dos Reis

Curso: Engenharia de Controle e Automação

Com a finalidade de estudar a expansão do universo, a cosmologia busca classificar diferentes tipos de objetos astronômicos. Entretanto, com o crescente aumento do número de objetos detectados, o método normalmente usado para a classificação (espectroscopia) torna-se muito custoso. Como consequência, utiliza-se um método com baixo custo (fotometria) embasado em algoritmos de aprendizado de máquina para a classificação desse vasto número de dados. Nesse contexto, o presente trabalho estuda otimizações para a melhoria desses algoritmos de aprendizado de máquina.

Abstract of Undergraduate Project presented to POLI/UFRJ as a partial fulfillment of the requirements for the degree of Engineer.

## SUPERNOVA PHOTOMETRIC IDENTIFICATION USING MACHINE LEARNING ALGORITHMS

Felipe Matheus Fernandes de Oliveira

July/2019

Advisors: Amit Bhaya

Ribamar Rondon de Rezende dos Reis

Course: Automation and Control Engineering

To explore the expansion history of the universe, cosmology classifies different types of astronomic objects using spectroscopy. With their sample sizes increasing, spectroscopy methods cannot handle this amount of data. As a solution to this issue, photometric identification is crucial to fully exploit these large samples due to its simplicity. Once photometric identification uses machine learning algorithms, the following work tries to optimize those algorithms.

# Sumário

<b>Lista de Figuras</b>	<b>x</b>
<b>Lista de Tabelas</b>	<b>xi</b>
<b>Lista de Símbolos</b>	<b>xii</b>
<b>Lista de Abreviaturas</b>	<b>xiii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Tema e Contextualização . . . . .	1
1.2 Problemática . . . . .	2
1.3 Delimitação . . . . .	3
1.4 Objetivos . . . . .	3
1.4.1 Processo Gaussiano <i>Gaussian Process</i> . . . . .	3
1.4.2 Aprendizagem Profunda . . . . .	4
1.4.3 Tratamento de <i>Outliers</i> . . . . .	4
1.5 Metodologia . . . . .	4
1.5.1 Desenvolvimento . . . . .	4
1.5.2 Avaliação dos Resultados . . . . .	4
1.6 Descrição . . . . .	4
<b>2 Fundamentos Teóricos</b>	<b>5</b>
2.1 Processo Gaussiano <i>Gaussian Process</i> . . . . .	5
2.2 Rede Neural Convolucional . . . . .	5
2.3 Demais Conceitos . . . . .	5
2.3.1 Análise de Componentes Principais . . . . .	5
2.3.2 Transformadas de <i>Wavelet</i> . . . . .	5
2.3.3 Validação Cruzada . . . . .	5
<b>3 Modelo Atual</b>	<b>6</b>
3.1 Pré-processamento dos Dados Brutos . . . . .	6
3.2 <i>Pipeline Atual</i> . . . . .	6



<b>4</b>	<b>Tratamento de <i>Outliers</i></b>	<b>7</b>
4.1	Estratégias Utilizadas . . . . .	7
4.2	Resultados e comparações . . . . .	7
<b>5</b>	<b>Rede Neural Convolucional</b>	<b>8</b>
5.1	Geração de Imagens e Parâmetros . . . . .	8
5.2	Escolhas de Redes Neurais . . . . .	8
5.3	Resultados e comparações . . . . .	8
<b>6</b>	<b>Interpolação através de Processo Gaussiano</b>	<b>9</b>
6.1	Escolha da Biblioteca . . . . .	9
6.2	Aleatoriedades e <i>Random Seeds</i> . . . . .	9
6.3	<i>Kernel Functions</i> . . . . .	9
6.4	Demais Observações . . . . .	9
6.5	Resultados das Interpolações . . . . .	9
6.6	Identificação e Justificativa do Erro . . . . .	9
<b>7</b>	<b>Resultados e Discussões</b>	<b>10</b>
<b>8</b>	<b>Conclusões</b>	<b>11</b>
8.1	Conclusões Finais . . . . .	11
8.2	Trabalhos Futuros . . . . .	11
<b>9</b>	<b>Introdução</b>	<b>12</b>
<b>10</b>	<b>Revisão Bibliográfica</b>	<b>14</b>
<b>11</b>	<b>Método Proposto</b>	<b>15</b>
<b>12</b>	<b>Resultados e Discussões</b>	<b>16</b>
<b>13</b>	<b>Conclusões</b>	<b>17</b>
	<b>Referências Bibliográficas</b>	<b>18</b>
<b>A</b>	<b>Algumas Demonstrações</b>	<b>19</b>

# Lista de Figuras

1.1	Exemplo da interpolação referente a um gráfico de fluxo de luz. . . .	2
9.1	Logotipo da POLI-UFRJ. . . . .	13
9.2	Logotipo da COPPE. . . . .	13

# Lista de Tabelas

9.1	Siglas dos cursos de engenharia da Escola Politécnica da UFRJ. . . .	12
9.2	Siglas dos programas de pós graduação da COPPE. . . . .	13
10.1	Exemplos de citações utilizando o comando padrão <code>\cite</code> do $\text{\LaTeX}$ e o comando <code>\citet</code> , fornecido pelo pacote <code>natbib</code> . . . . .	14

# Lista de Símbolos

$\emptyset$	Conjunto vazio, p. 1
$\mathbb{R}$	Conjunto dos números reais, p. 1

# Lista de Abreviaturas

COPPE	Instituto Alberto Luiz Coimbra de Pós-graduação e Pesquisa de Engenharia, p. 1
POLI-UFRJ	Escola Politécnica da Universidade Federal do Rio de Janeiro, p. 1

# Capítulo 1

## Introdução

### 1.1 Tema e Contextualização

Dentro da cosmologia, existe a necessidade de determinar distâncias luminosas ?? para modelar seus estudos, e como ferramenta dessa tarefa, utilizam-se as curvas de luz provenientes de supernovas do tipo IA.

No passado, o conjunto de dados de supernovas era pequeno o suficiente para poder analisar a maior parte dos objetos usando o método da espectroscopia, um método lento e custoso que, no entanto, confirma precisamente o tipo de cada uma delas.

Contudo, com o avanço das pesquisas e da tecnologia de telescópios, a astronomia está entrando em uma era de conjuntos massivos de dados, tornando-se necessário a adoção de técnicas automatizadas mais simples e práticas para classificar a enorme quantidade de objetos astronômicos captados, pois, através da espectroscopia não seria possível.

Nesse contexto, foram desenvolvidas diferentes abordagens para levantar essa grande quantidade de objetos captados através de um método chamado fotometria. Dentre essas abordagens, várias utilizam aprendizado de máquina.

Por fim, a ideia do projeto foi derivada de um artigo publicado pela cientista LOCHNER [1]. A autora busca criar uma maneira automática de classificação fotométrica usando as curvas de luz obtidas através da fotometria, essas que já foram devidamente classificadas no passado utilizando espectroscopia. Caracterizando assim um problema de classificação, visto que temos os tipos de supernovas dados pela espectroscopia, e suas representações fotométricas para que possamos criar e treinar modelos.

## 1.2 Problemática

Tendo em vista que o artigo [1] testou e validou diversos *pipelines*, neste trabalho adotamos aquele que ela obteve o melhor resultado de classificação, e a partir desse ponto, aplicamos modificações e avaliamos quaisquer possíveis melhoras.

O *pipeline* escolhido é constituído majoritariamente de 3 partes:

- Processo Gaussiano
- Transformada de Wavelet
- Floresta Aleatória

A problemática encontra-se no método de interpolação chamado **Processo Gaussiano**. Por ser um método de interpolação, espera-se que o mesmo defina uma função que passe pelos pontos obtidos respeitando suas incertezas. Entretanto, em alguns casos o gráfico interpolado é uma constante 1.1. O que não possui sentido físico, visto que se trata do fluxo de luz após uma explosão de um objeto astronômico.

Através do Instituto de Física da UFRJ surgiu a ideia principal deste projeto que é a de tentar corrigir essas interpolações que apresentam comportamento constante durante toda a observação.

Em paralelo, outras ideias foram aplicadas na tentativa de obter uma melhora no algoritmo original da *Lochner*.

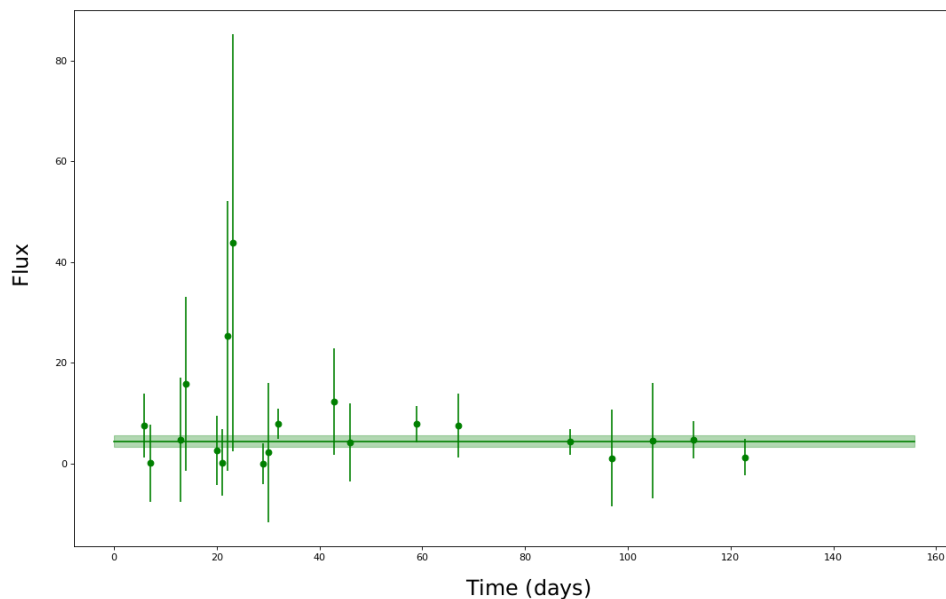


Figura 1.1: Exemplo da interpolação referente a um gráfico de fluxo de luz.

## 1.3 Delimitação

Todos os dados utilizados tanto neste trabalho quanto no artigo [1], foram extraídos da base de dados oferecida pelo desafio KESSLER [2]. A base de dados é de domínio público e pode ser obtida no [LINK DO DOWNLOAD](#).

Devido a granularidade dos dados brutos, é necessário um pré-tratamento visando buscar apenas os dados que iremos utilizar. Esse pré-tratamento foi aproveitado do *pipeline* do Instituto de Física da UFRJ, cujo processo visa ler os arquivos de texto transformando cada informação em uma chave de dicionário.

Ao fim do uso da seleção de dados citada acima, obtivemos valores em forma de dicionários em Python para cada objeto astronômico. Esses dicionários serão os **Dados Brutos** no escopo deste trabalho.

## 1.4 Objetivos

O objetivo do trabalho é otimizar determinados pontos do *pipeline* utilizado atualmente, após um estudo inicial ter sido feito para identificar pontos frágeis ou passíveis de melhoras.

Esses pontos são apresentados e justificados a seguir.

### 1.4.1 Processo Gaussiano *Gaussian Process*

O Processo Gaussiano (*Gaussian Process* ou **GP**), é a primeira parte do tratamento de dados após a transformação dos dados brutos em dicionários de Python.

Cada objeto astronômico irá possuir uma quantidade de pontos que representam a intensidade do fluxo de luz captado em determinado dia, bem como a incerteza desse valor. Assim, busca-se fazer uma interpolação para obter um gráfico que passe por determinados pontos.

O objetivo ao abordar esse ponto é consertar interpolações que não condizem com a interpretação física, como pode ser vista na figura 1.1



### 1.4.2 Aprendizagem Profunda

### 1.4.3 Tratamento de *Outliers*

## 1.5 Metodologia

### 1.5.1 Desenvolvimento

### 1.5.2 Avaliação dos Resultados

## 1.6 Descrição

Dizer aqui o que vai ter nos capítulos a frente

# Capítulo 2

## Fundamentos Teóricos

### 2.1 Processo Gaussiano *Gaussian Process*

### 2.2 Rede Neural Convolutacional

### 2.3 Demais Conceitos

Conceitos menos usados ao longo do *pipeline*, estão brevemente explicados a seguir.

#### 2.3.1 Análise de Componentes Principais

#### 2.3.2 Transformadas de *Wavelet*

#### 2.3.3 Validação Cruzada

# Capítulo 3

## Modelo Atual

### 3.1 Pré-processamento dos Dados Brutos

### 3.2 *Pipeline Atual*

# Capítulo 4

## Tratamento de *Outliers*

### 4.1 Estratégias Utilizadas

### 4.2 Resultados e comparações

# Capítulo 5

## Rede Neural Convolucional

A ideia foi blablabla usar deep learning, vimos exemplos das roupas, e colocamos lá deixamos a princípio o pipeline do Marcelo

### 5.1 Geração de Imagens e Parâmetros

Os parâmetros na hora de gerar as imagens e o PQ

### 5.2 Escolhas de Redes Neurais

escolhemos 2CNN pra fazer por isso e por aquilo

### 5.3 Resultados e comparações

# Capítulo 6

## Interpolação através de Processo Gaussiano

### 6.1 Escolha da Biblioteca

### 6.2 Aleatoriedades e *Random Seeds*

### 6.3 *Kernel Functions*

### 6.4 Demais Observações

Valor negativo e Mean function, erro grande, melhor overfittado que underfittado

### 6.5 Resultados das Interpolações

### 6.6 Identificação e Justificativa do Erro

## Capítulo 7

### Resultados e Discussões

# Capítulo 8

## Conclusões

### 8.1 Conclusões Finais

Falar que o Algoritmo de hoje já é bem robusto

### 8.2 Trabalhos Futuros

Falar de consertar o Erro e pah



# Capítulo 9

## Introdução

Segundo a norma de formatação de teses e dissertações do Instituto Alberto Luiz Coimbra de Pós-graduação e Pesquisa de Engenharia (COPPE), toda abreviatura deve ser definida antes de utilizada.

Do mesmo modo, é imprescindível definir os símbolos, tal como o conjunto dos números reais  $\mathbb{R}$  e o conjunto vazio  $\emptyset$ .

Você deve selecionar seu curso de engenharia usando o comando `\department{Sigla}` e no lugar de Sigla inserir a sigla referente ao seu curso de engenharia. A tabela 9.1 relaciona as siglas dos cursos de engenharia da Escola Politécnica da Universidade Federal do Rio de Janeiro (POLI-UFRJ), enquanto que a tabela 9.2 relaciona as siglas dos programas de pós graduação da COPPE.

Tabela 9.1: Siglas dos cursos de engenharia da Escola Politécnica da UFRJ.

Sigla	Curso
EA	Engenharia Ambiental
ECV	Engenharia Civil
ECI	Engenharia de Computação e Informação
ECA	Engenharia de Controle e Automação
EMAT	Engenharia de Materiais
EPT	Engenharia de Petróleo
EPR	Engenharia de Produção
EEC	Engenharia Eletrônica e de Computação
EET	Engenharia Elétrica
EMC	Engenharia Mecânica
EMET	Engenharia Metalúrgica
ENO	Engenharia Naval e Oceânica
ENU	Engenharia Nuclear

Note também que todas as figuras ou tabelas devem ser citadas no texto. Como ocorre com as tabelas 9.1 e 9.2. Para ilustrar o uso de figuras em  $\text{\LaTeX}$ , considere as figuras 9.1 e 9.2.

Tabela 9.2: Siglas dos programas de pós graduação da COPPE.

Sigla	Curso
PEB	Engenharia Biomédica
PEC	Engenharia Civil
PEE	Engenharia Elétrica
PEM	Engenharia Mecânica
PEMM	Engenharia Metalúrgica e de Materiais
PEN	Engenharia Nuclear
PENO	Engenharia Oceânica
PPE	Planejamento Energético
PEP	Engenharia de Produção
PEQ	Engenharia Química
PESC	Engenharia de Sistemas e Computação
PET	Engenharia de Transportes



Figura 9.1: Logotipo da POLI-UFRJ.



Figura 9.2: Logotipo da COPPE.

# Capítulo 10

## Revisão Bibliográfica

Para ilustrar a completa adesão ao estilo de citações e listagem de referências bibliográficas, a Tabela 10.1 apresenta citações de alguns dos trabalhos contidos na norma fornecida pela CPGP da COPPE, utilizando o estilo numérico.

Tabela 10.1: Exemplos de citações utilizando o comando padrão `\cite` do  $\text{\LaTeX}$  e o comando `\citet`, fornecido pelo pacote `natbib`.

Tipo da Publicação	<code>\cite</code>	<code>\citet</code>
Livro	[? ]	? ]
Artigo	[? ]	? ]
Relatório	[? ]	? ]
Relatório	[? ]	? ]
Anais de Congresso	[? ]	? ]
Séries	[? ]	? ]
Em Livro	[? ]	? ]
Dissertação de mestrado	[? ]	? ]
Tese de doutorado	[? ]	? ]

É importante notar que, segundo a Norma para a Elaboração Gráfica do Projeto de Graduação da Escola Politécnica da UFRJ para trabalhos de conclusão de curso de engenharia de julho de 2012, as referências bibliográficas podem ser apresentadas de duas formas: (i) Referências numeradas e (ii) Referências em ordem alfabética. Para exibição numerada, em que a exibição das referências bibliográficas segue a ordem de citação usada no texto, use o comando `\bibliographystyle{coppe-unsrt}`. Para exibição de referências bibliográficas em ordem alfabética, basta usar o comando `\bibliographystyle{coppe-plain}` ao final do documento.

# Capítulo 11

## Método Proposto

## Capítulo 12

### Resultados e Discussões

## Capítulo 13

### Conclusões

# Referências Bibliográficas

- [1] LOCHNER, M. “Photometric Supernova Classification with Machine Learning”, *ApJS (2016)*, v. 225, n. 2, pp. 31, set. 2016.
- [2] KESSLER, R. “Supernova Photometric Classification Challenge”, *arXiv*, v. 6, n. 5210, pp. 4, abr. 2010.

# Apêndice A

## Algumas Demonstrações