

Pós-Graduação Lato Sensu
Curso de Especialização em Inteligência Artificial

Introdução a Inteligência Artificial

Prof. Dr. Lucas Dias Hiera Sampaio

PARTE 05

As partes não podem ser compiladas integralmente.
Para uso exclusivo do curso de Pós Graduação da Universidade.



UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

CAMPUS CORNÉLIO PROCÓPIO

Frameworks para Inteligência Artificial

1. Apresentação

Na última semana, estudamos exemplos de diferentes bibliotecas e suas funções no contexto da inteligência artificial. Nesta semana vamos aprender um pouco mais sobre *frameworks* utilizados no contexto de Inteligência Artificial.

2. Frameworks

Frameworks são abstrações de software que provêm funcionalidade genérica para um determinado contexto e/ou aplicação. Diferentemente das bibliotecas, os frameworks são estruturas de código que contém as funções básicas para suportar uma determinada estrutura e fica a cargo do programador incluir as funções e informações complementares para que um objetivo seja atingido.

Neste sentido, *frameworks* vão além do código e dispõe como deve ser organizado dados, parâmetros e até mesmo a lógica/sequência de execução de modelos e ferramentas.

Atualmente existem diferentes *frameworks* que podem ser utilizados para prover soluções utilizando inteligência artificial sendo o mais conhecido deles o **Tensor Flow** que conta hoje com mais de 1500 colaboradores no projeto. Na próxima seção iremos abordar em detalhes aspectos do *framework*.

3. Tensor Flow

O TensorFlow é um *framework* livre e de código aberto para diferentes aplicações em inteligência artificial. O principal foco do TensorFlow é no processo de treinamento e inferência de redes neurais profundas (*Deep Neural Networks*).

Ele foi criado em 2015 pelo time da Google chamado Google Brain e hoje pode ser utilizado em diferentes linguagens de programação como Python, Javascript, C++ e Java.

Em paralelo ao desenvolvimento do framework a Google passa a desenvolver a partir de 2015 também as unidades de processamento Tensor (*Tensor Processing Unit*, TPU) que são circuitos integrados de aplicação específica (*application-specific integrated circuit*, ASIC) utilizados para acelerar o processamento de aplicações de inteligência artificial.

Ao longo dos anos, a Google continuou o processo de desenvolvimento deste hardware lançando em 2016 a primeira versão, em 2017 a segunda, 2018 a terceira e 2021 a quarta geração de TPUs.

Hoje as principais placas de vídeo da fabricante nVidia utilizam na sua arquitetura a quarta geração de tensor cores (são 512 tensor cores no modelo RTX 4090).

De forma geral, a maior parte das aplicações utilizando TensorFlow envolvem imagens e/ou vídeo que são artefatos compostos por uma grande quantidade de informações o que torna extremamente custoso o processamento em arquiteturas cujo processamento é sequencial. Por esta razão, as arquiteturas de hardware que são voltadas para IA tem como foco o desenvolvimento de soluções capazes de computação paralela.

Na seção seguinte, vamos investigar como utilizar de forma básica o TensorFlow dentro do contexto de redes neurais artificiais aplicadas em imagens.

4. Utilizando o TensorFlow

Suponha que seu objetivo seja criar uma IA capaz de classificar figuras em diferentes categorias, porém você não tem a menor ideia de como extrair informações destas imagens. Pois bem, é possível utilizar o TensorFlow para te auxiliar neste processo.

Contudo, para atingir esse objetivo, faz-se necessário estabelecer quais são os procedimentos, mesmo que de forma genérica, para se atingir o resultado esperado, isto é, a classificação das imagens.

A fim de auxiliar neste processo, a Figura 1 (na próxima página) ilustra um fluxograma com o passo a passo para resolver este problema. Note que a primeira etapa é o **pré-processamento**. Por exemplo, na videoaula, o conjunto de dados é composto por figuras de 28x28 pixels em escala de cinzas, isto é, cada pixel possui um valor entre 0 e 255. Para utilizar essas imagens no TensorFlow, o valor dos pixels é dividido pelo valor máximo, 255, de tal forma que os pixels passam a ter valor entre 0 e 1.

A segunda etapa é **separar o dataset** entre conjunto de treino e teste. Essa divisão é fundamental para garantir o processo de generalização da técnica de inteligência artificial. Utilizar o mesmo conjunto para treinamento e teste irá contaminar sua solução e consiste em grave falha na metodologia para solucionar o problema.

A terceira etapa é **construir a rede neural artificial**, ou seja, definir sua topologia - quantas camadas compõe a rede, qual o tamanho de cada camada, quais funções matemáticas são utilizadas, etc.

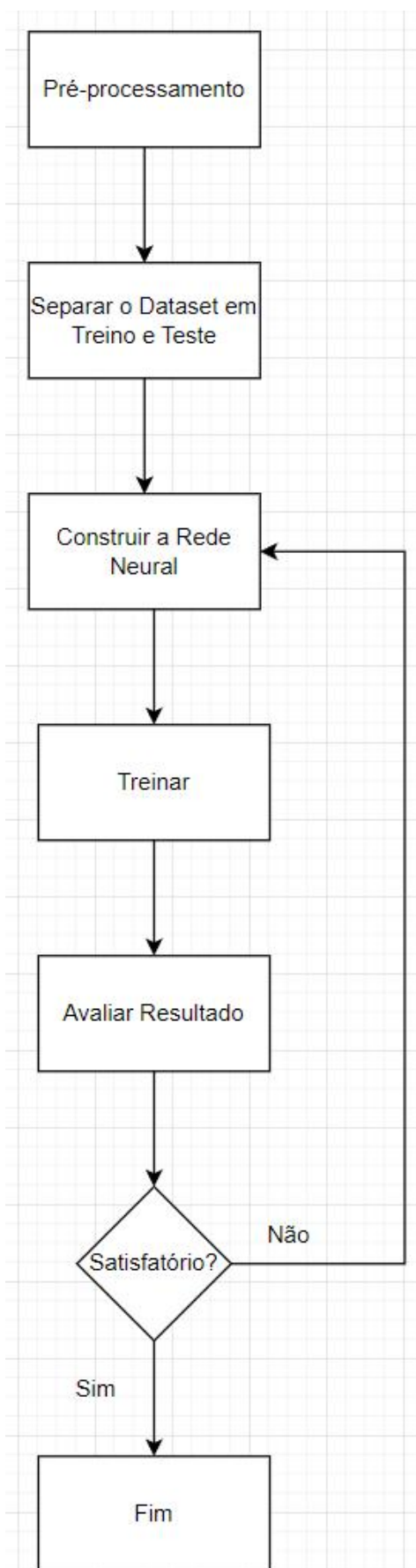


Figura 1: Fluxograma para realizar o treinamento e teste de uma rede neural para solucionar um problema de classificação.

O quinto passo é o **treinamento** em si. De forma geral, os frameworks possuem diferentes algoritmos que podem ser utilizados para realizar essa etapa e muitas vezes as diferenças, além do desempenho, estão na complexidade e quantidade de memória utilizada por estes algoritmos.

O sexto passo é **avaliar os resultados obtidos** que podem ser feitos por meio de métricas estatísticas como acurácia, especificidade, etc, ou por meio de funções específicas de perda que medem quão diferente a solução obtida pela rede neural é da solução esperada. Se o resultado for satisfatório de acordo com seus requisitos o processo é finalizado, caso contrário, deve alterar a topologia e parâmetros da rede neural para iniciar um novo processo de treinamento.

Na videoaula desta semana iremos realizar cada uma destas etapas e ver na prática os resultados obtidos por meio do uso do *framework* **TensorFlow**.