



9ª Jornada Científica e Tecnológica do IFSULDEMINAS

6º Simpósio da Pós-Graduação

ESTUDO DE TÉCNICAS DE APRENDIZAGEM DE MÁQUINA UTILIZANDO O SCIKIT-LEARN

Nykollas A. A. DIAS¹, Matheus E. FRANCO^{1,2}, João P. C. PAPP¹; Mateus R. BRUST¹;

RESUMO

Métodos de aprendizado profundo de máquina vêm se tornando gradualmente uma das técnicas mais utilizadas em diversos setores, devido a isso, é relevante para qualquer profissional e estudante da área de computação conhecer tais métodos e também as ferramentas que implementam esta tecnologia com o objetivo de garantir a sua competitividade no mercado e agregar valor aos seus serviços prestados. Desta maneira, foi realizado um levantamento das ferramentas e métodos que possam facilitar a implementação das técnicas de Aprendizagem de Máquina, além de mensurar sua aplicabilidade com estudantes da disciplina de Inteligência Artificial do IFSULDEMINAS – Campus Machado.

Palavras-chave: Inteligência Artificial, Machine Learning, Scikit-Learn.

1. INTRODUÇÃO

Atualmente, devido ao acelerado desenvolvimento da ciência, uma grande quantidade de informações é gerada dificultando a utilização das mesmas devido à inadequação dos sistemas convencionais de processamento de dados e a ineficácia do trabalho intelectual humano na execução dessa tarefa (CONDIE et al., 2013). Como solução foram desenvolvidas técnicas que permitem aos sistemas computacionais trabalharem com grandes volumes de dados aprendendo a executar tarefas sem serem explicitamente programados para isso. Elas são denominadas por técnicas de *Machine Learning* (ML) ou Aprendizagem de Máquina e representam um campo da Inteligência Artificial (IA) (CONDIE et al., 2013).

Dentre as técnicas de *Machine Learning*, existe um grupo de algoritmos capazes de aprender a partir de múltiplas camadas de processamento para solucionar problemas de classificação e de análise de padrões. Essas técnicas são denominadas por técnicas de *Deep Learning* (DENG; YU, 2014).

Najafabadi et al. (2015) afirmam que as técnicas de *Deep Learning* são utilizadas em aplicações de reconhecimento de fala, visão computacional, processamento de linguagem natural e apresentam alto desempenho nas suas implementações. Devido à alta complexidade dos modelos matemáticos que fundamentam a IA, faz-se necessária uma ferramenta que

¹IFSULDEMINAS – Campus Machado. ²Autor correspondente: matheus.franco@ifsuldeminas.edu.br

facilite a implementação desta tecnologia para que profissionais não experientes na área de IA possam utilizá-las de maneira criativa. Pensando nesta necessidade um grupo de pesquisadores em 2010 liderados por Gael Varoquaux (PEDREGOSA et al., 2011) desenvolveram uma biblioteca em código aberto chamada de scikit-learn. Esta biblioteca foi desenvolvida em Python e possui a propriedade de possibilitar o desenvolvimento de soluções de ML sofisticadas com poucas linhas de códigos além de se constituir de uma sintaxe extremamente simples. Hackeling (2014) afirma que o scikit-learn é bem documentado versátil e fácil de utilizar além de demandar apenas um conhecimento intermediário em programação. Portanto o scikit-learn se apresenta como uma ferramenta ideal para se democratizar o ML.

Este artigo apresenta os resultados parciais deste trabalho, cujos objetivos alcançados até o momento foi a implementação de uma técnica de Aprendizagem de Máquina utilizando-se o scikit-learn para classificar um banco de imagens de dígitos escritos à mão além de mensurar sua aplicabilidade de acordo com a visão de estudantes da disciplina de Inteligência Artificial do IFSULDEMINAS – Campus Machado.

2. MATERIAL E MÉTODOS

Inicialmente foram levantadas as ferramentas/bibliotecas que poderiam ser utilizadas ao propósito deste trabalho, listando-se Apache Mahout, Weka, TensorFlow e scikit-learn. A partir disto, definiu-se a utilização da biblioteca scikit-learn por sua simplicidade de implementação, preparação do ambiente e ampla utilização. Então foi implementado um método de aprendizado de máquina baseado em uma rede neural de múltiplas camadas MLP (*Multi Layer Perceptron*) por meio do scikit-learn 1.18 para classificar um conjunto de dados constituído de imagens de dígitos escritos à mão denominados por MNIST.

O objetivo desta foi classificar cada dígito de acordo com o padrão que eles aproximavam representar. Uma parte do conjunto de dados foi utilizada no processo de treinamento do classificador, e a parte restante foi utilizada no processo de validação de seu treinamento. Foram utilizadas como conjunto de dados 1797 imagens de 8 x 8 pixels, armazenadas em uma matriz com as respectivas dimensões 1797 x 8 x 8.

Durante o processo de treinamento o programa iterou cada imagem incluída no conjunto de treinamento associando cada unidade a seu respectivo valor. Após a conclusão deste processo foi executada a validação do treinamento utilizando-se as imagens incluídas no outro conjunto. Durante a validação o programa criou uma lista constituída pelas imagens e seus respectivos valores correspondentes.

Concluindo-se os processos o programa apresentou as imagens com a respectiva classificação realizada pelo classificador na parte superior. O trabalho foi apresentado para alunos do 7º período de Licenciatura em Computação, e foi realizado um levantamento com o fim de avaliar a experiência dos estudantes com a ferramenta. A pesquisa foi realizada com 12 estudantes que responderam seis perguntas relativas à experiência deles e relativas à avaliação deles no que tange a compreensibilidade.

3. RESULTADOS E DISCUSSÕES

Durante a impressão das imagens com suas respectivas classificações percebeu-se que a rede neural obteve êxito em classificar a maioria das imagens, acertando 93 %. A impressão converteu as matrizes 8x8 em imagens de forma que cada elemento da matriz recebeu uma intensidade de cor diferente de acordo com o valor contido no respectivo elemento. Os elementos da matriz que continham valor equivalente a zero foram representados pela cor branca, e os elementos que possuíam um valor maior que zero foram representados pela cor preta de intensidade diretamente proporcional aos seus valores.

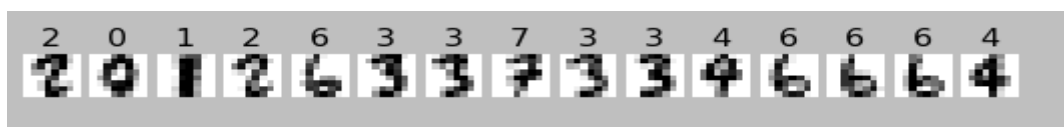


Figura 1: Resultado da classificação pela rede neural artificial.

Fonte: Autoria Própria

Os valores na parte superior representam a classificação realizada pela rede neural artificial. As imagens na parte inferior representam os exemplares do conjunto MNIST que foram classificados.

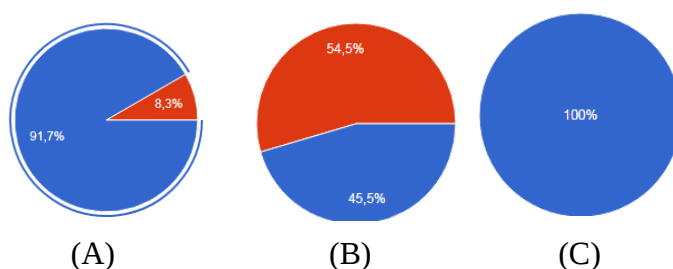


Figura 2 – Gráficos de resultados da pesquisa realizada com os alunos

Fonte: Autoria Própria

A pesquisa realizada com os estudantes do 7º período de Licenciatura da Computação mostrou que 82% consideram de interesse a aplicação do scikit-learn nos estudos da disciplina de IA. E também foi constatado que 91,7% (Figura 2A) dos estudantes

conseguiram perceber aplicações práticas para a ferramenta.

A pesquisa mostrou que 54,5% (Figura 2B) não conhecem outra ferramenta que apresenta a solução fornecida pelo scikit-learn e também que 73% dos estudantes sentem se motivados em implementar soluções de Machine Learning em diversos problemas, o que pode ser um indicativo de interesse por parte dos estudantes na área em geral.

Estas duas últimas informações sugerem que apesar de existir certo interesse por parte dos estudantes em aplicar soluções de Machine Learning, eles carecem de informações sobre ferramentas que possam tornar práticas tais aplicações.

A maioria dos estudantes afirmou que entenderam todas as funcionalidades apresentadas do scikit-learn. E constatou-se que 100 % (Figura 2C) consideram que o scikit-learn pode facilitar seus estudos sobre IA e suas aplicações de ML.

4. CONSIDERAÇÕES FINAIS

A utilização de algoritmos para treinar classificadores utilizando a biblioteca scikit-learn pôde ser corretamente implementada, possibilitando assim a compreensão sobre uma ferramenta de Aprendizagem de Máquina para identificação de padrões. Os resultados apontam que é possível obter saídas confiáveis com essa técnica. Além disto, pode-se divulgar a técnica para estudantes de IA para que além de obter o conhecimento sobre o assunto, possam desenvolver soluções baseadas nesta tecnologia. Este trabalho está em desenvolvimento para o estudo de novas técnicas e suas aplicações.

REFERÊNCIAS

- CONDIE, T. *et al.* Machine learning on big data. 2013, [S.l.]: IEEE, 2013. p. 1242–1244.
- DENG, L.; YU, D. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, v. 7, n. 3–4, p. 197–387, 2014.
- HACKELING, G. *Mastering Machine Learning with scikit-learn*. [S.l.]: Packt Publishing Ltd, 2014.
- NAJAFABADI, M. M. *et al.* Deep learning applications and challenges in big data analytics. *Journal of Big Data*, v. 2, n. 1, p. 1, 2015.
- PEDREGOSA, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, n. Oct, p. 2825–2830, 2011.