

UNIVERSIDADE FEEVALE

BRUNO MACIEL DE FARIAS

REDES NEURAIS PARA
CLASSIFICAÇÃO AUTOMÁTICA DE DOCUMENTOS
(Título Provisório)

Anteprojeto de Trabalho de Conclusão

Novo Hamburgo
2021

BRUNO MACIEL DE FARIAS

REDES NEURAIIS PARA
CLASSIFICAÇÃO AUTOMÁTICA DE DOCUMENTOS

(Título Provisório)

Anteprojeto de Trabalho de Conclusão de
Curso, apresentado como requisito parcial
à obtenção do grau de Bacharel em
Ciência da Computação pela
Universidade Feevale

Orientador: Gabriel da Silva Simões

Novo Hamburgo
2021

RESUMO

Com a evolução dos algoritmos e com o aumento da capacidade de processamento por computadores, a implementação de tecnologias informacionais baseadas no paradigma de inteligência artificial se torna viável e vem ganhando grande destaque no meio organizacional. Essas tecnologias possibilitam a construção de soluções efetivas para os mais variados tipos de problemas, inclusive para o tema de classificação de imagens, mais especificamente classificação de Documentos Oficiais, tema que será abordado como principal problemática a ser resolvida neste trabalho. O presente trabalho apresenta uma fundamentação teórica sobre aprendizado profundo e o funcionamento de redes neurais convolucionais, que servirão como base para o desenvolvimento de um protótipo para experimentação. Os resultados obtidos serão comparados com outros meios de classificação, e apresentados na sequência.

Palavras-chave: *Machine Learning*; Aprendizado de Máquina; Redes Neurais Convolucionais; identificação de documentos; classificação de documentos.

SUMÁRIO

MOTIVAÇÃO	5
OBJETIVOS	8
METODOLOGIA	9
CRONOGRAMA	10
BIBLIOGRAFIA	11

MOTIVAÇÃO

A necessidade de expansão para novos mercados fez com que uma empresa com 45 anos de atuação no ramo de varejo automobilístico no sul do Brasil enxergasse a área de produtos financeiros como uma grande oportunidade para expandir seus negócios nacionalmente. Com essa premissa, a empresa descreve o procedimento de solicitação de imagens de documentos para o cadastro de clientes da seguinte maneira:

O funcionário solicita que o cliente encaminhe imagens de seus documentos via e-mail, WhatsApp ou de forma física. Os dados são recebidos e armazenados em uma base de dados, onde a realização de uma inspeção visual somente ocorre caso o usuário tenha de processar os dados do cliente naquele exato momento.

Sabe-se que a implementação de tecnologias automatizadas em substituição a procedimentos e mecanismos manuais é uma das chaves que distingue empresas de sucesso (MATTOS, 2002). Portanto o trabalho busca implementar tecnologias que aplicam conceitos de Inteligência Artificial (IA), para otimizar o processo de reconhecimento de documentos, demanda apresentada pela empresa.

Com o risco da simplificação, podemos definir IA como o desejo de atribuir características antropomorfizadas em autômatos, a fim de que estes possam tomar decisões baseadas no livre-arbítrio. Esta característica permite que o agente desenvolva um método intrínseco a sua própria “cognição” para realizar tarefas ou apresentar soluções. Tal conceito vem sendo idealizado desde a longínqua década de 1940 (DONEDA, 2018).

Sendo uma área de ciência pouco delimitada, a IA versa com as áreas de conhecimento, proporcionando que diversos pesquisadores utilizem de suas benesses. Wooldridge (2019) se refere a IA como um pó de fadas, composto que quando aplicado sobre alguma superfície ou objeto lhe proporciona poderes mágicos. Apesar da similaridade, pode-se dizer com certeza que a parte mágica fica por conta da uma poderosa fundamentação matemática, que possibilita o desenvolvimento de modelos numéricos capazes de transformar grandes quantidades de dados em informações, onde em razão do esforço necessário seriam humanamente inalcançáveis. A vasta utilização destes conceitos populariza e traz o termo para as manchetes de periódicos e revistas mundiais corriqueiramente (WEEKS, 1996; WOOLDRIDGE, 2019).

Um dos campos de pesquisa de IA que vem ganhando notoriedade, é o de Visão Computacional (VC). Sendo a ciência que pretende proporcionar a capacidade de visão humana para as máquinas, VC é uma complexa área do conhecimento científico que extrai e identifica informações a partir de imagens observadas, gerando assim uma emulação da visão humana

(MAREGONI, 2010). Os recursos realizados para que a visão de fato aconteça vão desde o processamento até o reconhecimento e tipificação de informações presentes na cena analisada.

Segundo Weeks (1996 apud RUDEK, COELHO e CANCEGLIERI 2001), existe uma sequência de pontos essenciais para o processamento digital de imagem que são utilizados em VC, sendo estes:

i. Aquisição de Imagem: Compreende desde o hardware utilizado para obtenção da imagem, assim como hardware e software que a processam após a coleta;

ii. Pré-processamento: Procedimento que realiza correções na imagem, assim como ajustes de iluminação, contraste, distorções e nitidez.

iii. Segmentação: É o processo que realiza a divisão da imagem em regiões estruturais com o intuito de minimizar a informação contida na cena. Tal processo é de suma importância na otimização de performance quando os modelos realizam a identificação de objetos e reconhecimento de padrões;

iv. Identificação de Objeto: Visa encontrar e extrair objetos de uma determinada cena, apontando regiões mais significativas em uma imagem. Tal recurso possibilita que ações futuras sejam realizadas com mais precisão, restringindo o tamanho da imagem a ser analisada.

v. Reconhecimento de Padrões: Algoritmos de reconhecimento de padrões atuam com as informações extraídas da imagem, utilizando técnicas de extração de atributos como detecção de bordas e cantos, buscam encontrar semelhanças entre os objetos presentes na imagem para classificá-los. Ferramentas matemáticas computacionais são os maiores aliados dos engenheiros de VC nesta etapa, já que permitem o processamento e a comparação das imagens.

A natural singularidade de imagens obtidas para aplicação de VC é o principal fator de heterogeneidade quando se adentra em questões técnicas da ciência. Fatores externos como contraste, posicionamento, ângulo e ruídos gerais, aumentam a complexidade para uma correta interpretação da cena (RUDEK; COELHO; CANCEGLIERI, 2001). Devido a esta característica, aplicar a mesma técnica em diferentes cenários não gera bons resultados, pois imagens diferentes possuem requisitos de programação diferentes (MAHONY, 2019). Na busca pela otimização da execução destes procedimentos, o refinamento dos algoritmos, técnica conhecida como *Tunning*, faz a interferência manual de um engenheiro de VC ser vital (MAREGONI, 2010). Todavia, este procedimento vai de encontro com a demanda que ambientes empresariais exigem, onde a otimização de tempo e produtividade determinam quem

será o melhor e obterá as melhores remunerações (PISTONO, p.13). A fim de reduzir a interação do homem com a máquina, conceitos de aprendizado de máquina e aprendizado profundo serão apresentados.

Mitchell (1997) diz que sistemas que lidam com um volume exponencial de dados não estruturados são capazes de encontrar regras e correlações para fazer previsões e até mesmo tomar decisões para um determinado problema. Esta capacidade é atingida sem que os sistemas tenham sido explicitamente programados para isso, sendo então considerados como detentores de agentes inteligentes em seu algoritmo. A área da ciência que estuda este paradigma é denominada *Machine Learning* ou a Aprendizado de Máquina (ML).

Ao sintetizar o que seria “aprender” para uma máquina, Mitchell (1997) define da seguinte maneira, “um programa de computador aprende através de uma experiência E a respeito de um conjunto de tarefas T mensuráveis por uma métrica P se o seu desempenho na tarefa T, medida pela métrica P melhora com a experiência.”

ML trabalha construindo modelos matemáticos capazes de gerar previsões que podem contribuir para a tomada de decisões. Esta característica só é possível pois uma determinada amostra de dados (dados de treinamento) é disponibilizada para um modelo. Estes dados são divididos em duas partes, chamados de conjunto de treinamento e o conjunto de teste, onde o primeiro será utilizado para, de fato, treinar o modelo, enquanto o outro servirá como base para realização de testes e aferição da qualidade obtida. É importante ressaltar que dentro do conjunto de treinamento ainda existe mais uma importante divisão que se refere aos conjuntos de dados como rotulados e não rotulados. Dados rotulados são os aqueles que já possuem prévio conhecimento de sua classificação ou resultados, e não rotulados que não possuem nenhum vínculo ou resultado conhecido (MITCHELL, 2020)

As técnicas de ML são tradicionalmente divididas em aprendizado por reforço, aprendizado supervisionado e um terceiro denominado aprendizado não supervisionado. Tais paradigmas possuem as seguintes definições:

Aprendizado por Reforço: Com este método, a máquina atua sobre os dados cumprindo uma determinada tarefa, inicialmente com uma abordagem iterativa de tentativa e erro. Posteriormente, os resultados de cada tentativa, independentemente de seu sucesso, são anotados e utilizados para treinar o agente por meio de um sistema de recompensa/punição, que visa maximizar as recompensas futuras ao longo de sua execução, proporcionando melhores resultados (RUSSELL; NORVIG, 1995).

Aprendizado Supervisionado: É a abordagem que consiste em aprender a executar uma tarefa utilizando exemplos conhecidos, ou seja, utilizando dados rotulados (RUSSELL; NORVIG, 1995). Podemos exemplificar observando o histórico de vendas de um supermercado onde valores, itens e período de venda influenciam na predição de uma próxima promoção a ser criada. A abordagem supervisionada consiste em um método iterativo para obtenção de uma porcentagem aceitável de acertos. Como é conhecido o resultado, é possível que o algoritmo execute iterações quantas vezes forem necessárias, a fim de buscar a acurácia máxima que pode ser alcançada (RUSSELL; NORVIG, 1995).

Aprendizado Não-Supervisionado: É o paradigma que lida com dados não estruturados, onde dados e resultados esperados não são rotulados e conhecidos. Nessa implementação o algoritmo não receberá recompensa ao executar iterações, sendo chave desta abordagem o agrupamento. Esta tarefa trabalha agrupando os dados, e utilizando os métodos numéricos para buscar nas características e padrões de semelhanças, correspondências que possam deduzir algum sentido a estes (GHAHRAMANI, 2004) (RUSSELL; NORVIG, 1995).

Sabendo que técnicas de ML trabalham com características distintas em seus treinamentos, a fim de encontrar correlações em bases de dados cada vez mais gigantescas, pesquisadoras passaram a reunir conjuntos de algoritmos capazes de modelar abstrações em alto nível (ZHANG, 2020). A arquitetura destes algoritmos foi pensada de forma que camadas de processamento não-linear são utilizadas para a extração e transformação de características dos dados, onde cada camada sucessiva usa a saída da camada anterior como entrada. Uma característica interessante desta nova proposta é que estas camadas não são projetadas por engenheiros humanos, mas sim pelo próprio processo de aprendizagem. Por sua vez, esse novo paradigma recebeu o nome de *Deep Learning* (DL), ou Aprendizado Profundo (MENDONÇA, 2009), (ZHANG, 2020), (LECUN, 2015).

Dentro de DL encontram-se as Redes Neurais Artificiais (RNA's), que baseadas na neurociência, buscam simular computacionalmente, de maneira simplificada, padrões de comunicação e relações neuronais que acontecem no sistema nervoso (KRICHEVSKY, 2012). Mattos (2002) e Zhang (2020) definem que redes neurais são grupos de neurônios, nós de computação que transmitem valores entre si. Em um primeiro momento, o neurônio avalia qual o peso de um determinado valor de entrada, e aplica nele uma função que pode ser linear ou não-linear. O resultado desta operação será encaminhado para a próxima camada de neurônios realizar a mesma operação até que o modelo encontre o peso ideal para o dado que está sendo processado. Esse procedimento é conhecido como Treinamento.

Uma implementação que vem ganhando destaque devido a uma alteração na arquitetura das RNA's, são as Redes Neurais Convolucionais (RNC's). São implementações baseadas na arquitetura de uma RNA, que adicionam pelo menos uma camada de convolução ao modelo. Essa nova camada, faz uso de *Kernels* ou filtros, que contém uma matriz de valores (pesos), obtidos nas camadas anteriores de saídas dos neurônios. A função desse *Kernel* é fazer a convolução dos dados de entrada, visando encontrar possíveis relações nas matrizes da rede, permitindo que aconteça o reconhecimento de elementos da imagem (DONEDA, 2018) (O'MAHONY, 2019).

Recentes pesquisas apontam que a performance no reconhecimento de imagens por humanos é ligeiramente melhor do que modelos baseados em aprendizado profundo (DODGE, 2017) (MORGANTI, 1999). No entanto, estas pesquisas não consideram fatores externos como o estresse. Cruz (2020, p. 7), cita estudos que apontam que aproximadamente 72% dos trabalhadores brasileiros se sentem estressados com o seu trabalho. Dias (2008) relata que o processo de validar imagens é cansativo, demorado e tedioso, o que corrobora no surgimento de episódios de estresse. Dias (2008) ainda relaciona questões de estresse e pressão, com o desenvolvimento de implicações físicas e psicológicas sobre o profissional, incidindo diretamente na sua assiduidade, motivação, criatividade, produtividade e até mesmo levando-os a cometerem erros em seus julgamentos.

Este trabalho assume que a implementação de novas tecnologias faz parte de um conjunto de iniciativas que se encontra presente em empresas de sucesso, onde estas ações influenciam diretamente no bem-estar dos funcionários que, ao se depararem com rotinas automatizadas, passam a trabalhar de maneira mais efetiva, podendo dar atenção a tarefas mais produtivas. Portanto, considerando as premissas apresentadas, este trabalho assume também que a implementação de modelos de inteligência artificial baseados em redes neurais convolucionais pode solucionar os problemas relatados pela empresa no campo de validação de imagens de Documentos Oficiais. Outro significativo ponto que este trabalho proporcionará, pós a implantação de modelos preditivos, será a rotulagem automática dos dados que serão inseridos na base da empresa. Esta contribuição permitirá à empresa inúmeras facilidades no manuseio e no processamento de documentos, abrindo espaço para oportunidades futuras, especialmente na otimização de recursos e capital.

OBJETIVOS

Objetivo geral

Implementar e treinar e avaliar modelos de redes neurais convolucionais que permitam a classificação automática de documentos de acordo com um número específico de classes (inicialmente 4).

Objetivos específicos

- Realizar pesquisa bibliográfica sobre redes neurais convolucionais e classificação de imagens digitais;
- Implementar um modelo de redes neurais convolucionais;
- Treinar o modelo implementado;
- Implementar um protótipo para avaliação do modelo treinado;
- Avaliar o desempenho do modelo neural utilizado no protótipo;

METODOLOGIA

Segundo Prodanov e Freitas (2013), a pesquisa é caracterizada como de natureza aplicada, pois pretende solucionar um problema específico, que é a implementação de um modelo de Inteligência Artificial para otimizar a realização de classificação de modelos de Documentos Oficiais.

Utilizando o caráter de pesquisa bibliográfica, serão concisamente apresentados trabalhos correlatos que utilizam outros métodos de classificação de imagens com IA. Esta tarefa pretende endossar a necessidade/validade de implementação de modelos automatizados para a problemática especificada. Ainda com caráter bibliográfico, serão apresentados conceitos de Redes Neurais Convolucionais de maneira geral e aplicados na classificação de imagens.

Atendendo ao caráter de pesquisa experimental, serão investigadas:

- i. Técnicas de modelagem, a fim de serem aplicadas para definir como e quais algoritmos farão parte da composição do protótipo. A partir destas técnicas será possível definir a arquitetura para a concepção de um protótipo;
- ii. A construção e/ou avaliação de quais frameworks possuem as características necessárias para criação de um protótipo que valide os objetivos do trabalho com as tecnologias estudadas.

Como será necessário à validação do modelo, será realizado uma busca por uma ou mais bases de dados que sustentem o treinamento e o processo de validação dos resultados do estudo.

Por fim, com a prototipagem concluída, serão verificados os resultados obtidos com o modelo proposto comparando os valores preditos com os rótulos originais de cada instância. A partir destas comparações será possível a validação da otimização que as tecnologias implementadas podem oferecer ao ambiente empresarial.

CRONOGRAMA

Trabalho de Conclusão I

Etapa	Meses			
	Mar	Abr	Mai	Jun
Análise contextual sobre classificação de imagens	x			
Desenvolvimento do anteprojeto	x	x		
Entrega do anteprojeto		x		
Pesquisa sobre conceitos relacionados		x		
Pesquisa e definição de modelos para implementação		x	x	
Redação TCC I		x	x	x
Definir modelo de implementação				x
Entrega TCC I				x

Trabalho de Conclusão II

Etapa	Meses			
	Ago	Set	Out	Nov
Procura por bases de dados	x			
Implementação do modelo	x	x		
Análise dos resultados		x		
Comparação dos resultados		x	x	
Comparação entre o modelo proposto com o atual			x	
Escrita do TCC II		x	x	x
Entrega do TCC II				x

BIBLIOGRAFIA

DA CRUZ, Arthur Conrado Araújo. **Combustível: uma reflexão prático-teórica sobre o burnout universitário em formato audiovisual**, p. 1-388–416.

DIAS, SHEILA MARA OLIVEIRA. **Fatores de pressão no trabalho e comprometimento com a carreira: um estudo com profissionais de Tecnologia da Informação (TI)**. 2008. Tese de Doutorado. Dissertação de Mestrado em Administração da Faculdade de Ciências Econômicas da Universidade FUMEC. Pags. 170.

DODGE, Samuel; KARAM, Lina. **A study and comparison of human and deep learning recognition performance under visual distortions**. In: 2017 26th international conference on computer communication and networks (ICCCN). IEEE, 2017. p. 1-7.

DONEDA, Danilo Cesar Maganhoto et al. **Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal**. Pensar-Revista de Ciências Jurídicas, v. 23, n. 4, p. 1-17, 2018.

GHAHRAMANI, Zoubin. **Unsupervised learning**. In: Summer School on Machine Learning. Springer, Berlin, Heidelberg, 2003. p. 72-112. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-540-28650-9_5>. Acesso em: 20 mar. 2021

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. **Imagenet classification with deep convolutional neural networks**. Advances in neural information processing systems, v. 25, p. 1097-1105, 2012.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. **Convolutional networks and applications in vision**. In: IEEE International Symposium on Circuits and Systems. Paris, 2010. Disponível em: <https://www.researchgate.net/publication/221376179_Convolutional_Networks_and_Applications_in_Vision>. Acesso em: 12 mar. 2020.

MARENGONI, M; STRINGHINI, D. **Tutorial: Introdução à Visão Computacional usando OpenCV**. Porto Alegre, Universidade Presbiteriana Mackenzie, 2009. Disponível em: <<https://seer.ufrgs.br/rita/article/view/11477>>. Acesso em: 01 abr. 2021.

MATTOS, Lissandra Kerppers. **As mudanças organizacionais e seus gestores nas empresas na era da informação**. 2002. Tese de Doutorado. Disponível em: <<http://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/3671/Lissandra.pdf?sequence=1>>. Acesso em: 30 mar 2021.

MENDONÇA, Marco Aurélio Alves de; FREITAS, Fernando de Almeida; SOUZA, Jano Moreira de. **Tecnologia da informação e produtividade na indústria brasileira**. Revista de Administração de Empresas, v. 49, n. 1, p. 74-85, 2009.

MITCHELL, Tom M. et al. **Machine learning**. 1997.

MORGANTI, M. et al. **Automatic PCB inspection Algorithms**. Computer Vision and Image Understanding, v. 63, n. 2, p. 287-313, 1999.

O'MAHONY, Niall et al. **Deep learning vs. traditional computer vision**. In: Science and Information Conference. Springer, Cham, 2019. p. 128-144.

PISTONO, Federico. **Os robôs vão roubar seu trabalho, mas tudo bem: Como sobreviver ao colapso econômico e ser feliz**. Trad. Pedro Maia Soares. São Paulo: Portfolio-Penguin, 2017 p.13

RUDEK, Marcelo; COELHO, Leandro dos Santos; CANCEGLIERI JR, O. **Visão Computacional Aplicada a Sistemas Produtivos: Fundamentos e Estudo de Caso**. XXI Encontro Nacional de Engenharia de Produção-2001, Salvador, 2001.

RUSSELL, Stuart; NORVIG, Peter. **Artificial intelligence: a modern approach**. Nova Jersey, Prentice Hall, 2002. v. 2, cap. 6, p. 523- 635.

WEEKS, Arthur R. **Fundamentals of electronic image processing**. SPIE Optical Engineering Press, 1996.

WOOLDRIDGE, Michael. **Artificial Intelligence requires more than deep learning — but what, exactly?** Oxford: Elsevier, 2020. Disponível em: <<https://www.sciencedirect.com/sdfe/pdf/download/eid/1-s2.0-S0004370220301363/first-page-pdf>>. Acesso em: 10 mar. 2021.

YAN, Le Cun; YOSHUA, B.; GEOFFREY, H. **Deep learning**. Nature, v. 521, n. 7553, p. 436-444, 2015.

ZHANG, Xian-Da. **A Matrix Algebra Approach to Artificial Intelligence**. Springer, cap. 6, p. 223-440, 2020

ZHANG, Ziwei; CUI, Peng; ZHU, Wenwu. **Deep learning on graphs: A survey**. IEEE Transactions on Knowledge and Data Engineering, 2020.