

Aprendizado profundo na detecção contextual

Deep learning in contextual detection

RESUMO

Com a criação e popularização das redes sociais, o termo engajamento tornou-se importante para as grandes marcas. Além disso, o volume de dados gerado por todo planeta vem crescendo de maneira exponencial, estando em diferentes formatos, como imagens. Sabendo disso, o objetivo dessa pesquisa foi desenvolver uma rede neural convolucional (CNN), capaz de aprender e classificar imagens, julgadas interessantes e desinteressantes para a rede social Instagram. Para corroborar tal proposta, foram realizados experimentos com o nicho *Dog* e *Food*. Para tanto, 179400 imagens do nicho *Dog* e 120353 imagens do nicho *Food*, pertencentes à usuários do Instagram, foram coletadas, e após o pré-processamento, junto com outras informações como o número de curtidas e de seguidores, criou-se uma CNN que conseguisse classificar uma nova imagem nos nichos escolhidos. A acurácia total do nicho *Dog* foi de 91%, e a acurácia total do nicho *Food* foi de 82%, resultado esse considerado satisfatório, com potencial de ser ampliado se novas técnicas forem aplicadas, como utilizar outro tipo de análise na seleção de imagens interessantes e desinteressantes.

PALAVRAS-CHAVE: Inteligência artificial. Processamento de imagens. Redes neurais (Computação).

ABSTRACT

With the creation and popularization of social networks, the term engagement has become important for major brands. In addition, the volume of data generated by the entire planet has been growing exponentially, being in different formats, such as images. Knowing this, the objective of this research was to develop a convolutional neural network (CNN), capable of learning and classifying images, deemed interesting and uninteresting for the social network Instagram. To corroborate this proposal, experiments were carried out with the Dog and Food niche. For this, 179400 images from the Dog niche and 120353 images from the Food niche, belonging to Instagram users, were collected, and after pre-processing, along with other information such as the number of likes and followers, a CNN was created that managed to classify a new image in the chosen niches. The total accuracy of the Dog niche was 91%, and the total accuracy of the Food niche was 82%, a result considered satisfactory, with the potential to be expanded if new techniques are applied, such as using another type of analysis in the selection of interesting and uninteresting images.

KEYWORDS: Artificial intelligence. Image processing. Neural networks (Computation).

Vitor Fabrile Guastala
vitorguastala@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

Pedro Henrique Bugatti
pbugatti@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

Recebido: 19 ago. 2020.

Aprovado: 01 out. 2020.

Direito autoral: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



INTRODUÇÃO

Com a criação e popularização das redes sociais, o termo engajamento tornou-se importante para as grandes marcas. Esse termo refere-se a uma maneira de entender como o público interage com as marcas, a partir de curtidas, comentários, compartilhamentos, dentre outras métricas utilizadas para avaliar a ligação entre público e marcas nas redes sociais (ROCHA, 2019; SBCOACHING 2019).

Além disso, o volume de dados gerado por todo planeta vem crescendo de maneira exponencial, e, tendo em mente que os meios mais contribuintes para a geração desses dados são as redes sociais, cerca de 80% desses dados são não-estruturados, ou seja, não possuem uma estrutura pré-definida, ou estão em diferentes formatos, como é o exemplo das imagens (TAKAHASHI, 2015; MENDES, 2016).

Assim, o conceito de Big Data surgiu com a necessidade de interpretar e analisar esse grande conjunto de dados, e unido a técnicas de recomendações oriundas do *deep learning*, ou, aprendizado profundo, empresas são capazes de fazer com que informações cheguem aos seus clientes sem que ele necessite buscar por essas informações (TAKASHI, 2015; BUYYA; CALHEIROS; DASTJERDI, 2016; SANTANA, 2018).

Dessa maneira, por serem mais atrativas do que textos convencionais, as imagens são importantes na influência das escolhas dos usuários sobre as recomendações, logo, analisá-las traz benefícios no engajamento (SANTANA, 2018; MELO, 2016).

Uma rede neural convolucional (*convolutional neural network*, CNN) é um algoritmo de aprendizado profundo que utiliza uma rede neural para reconhecimento e classificação de imagens. Ela recebe como entrada uma imagem, que é enxergada pelo computador por uma matriz de pixels, e então, se treinada corretamente, é capaz de classificá-la em categorias como: cães, gatos, objetos, pessoas, dentre outros (PRABHU, 2018).

Sabendo disso, o objetivo dessa pesquisa foi desenvolver uma CNN capaz de aprender e classificar imagens julgadas interessantes e desinteressantes para a rede social Instagram em um determinado nicho.

MATERIAL E MÉTODOS

Para a execução desta pesquisa, foram utilizadas as seguintes tecnologias: LevPasha Instagram API, MongoDB, linguagem de programação Python, e bibliotecas Keras, Scikit-Learn e Matplotlib.

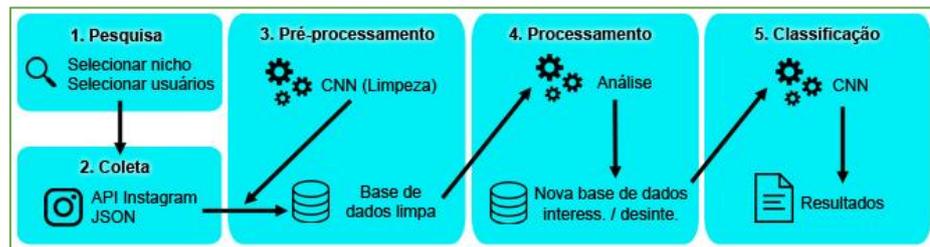
LevPasha Instagram API é uma API *open source* escrita em Python. Ela foi utilizada para obter informações como, o número de seguidores de cada usuário, as imagens publicadas, o número de curtidas e de comentários de cada publicação, dentre outros. Foi optado por utilizar esta API, pois a API oficial disponibilizada pelo Instagram era muito limitada se comparado a esta.

Apesar da API coletar todas as informações em JSON (uma formatação leve de troca de dados), foi necessário utilizar o MongoDB, um SGBD orientado a documentos, para armazenar os dados provindas da etapa de processamento.

Todo o projeto foi realizado com a linguagem de programação Python, junto com bibliotecas úteis relacionadas a IA e aprendizado profundo, sendo elas a Keras para a criação de redes neurais, a Scikit-Learn para a classificação e Matplotlib para melhor visualização dos resultados.

A Figura 1 mostra um esquema em que se resume todo o método aplicado neste estudo, desde a coleta de dados, até a visualização dos resultados.

Figura 1 – Esquema de todo método utilizado na pesquisa



Fonte: Autoria Própria (2020).

Para a execução desta pesquisa, foi necessário **escolher nichos** para serem trabalhados, mas, apesar da escolha de um nicho ser subjetiva, foi realizado uma pesquisa inicial que trouxesse nichos relevantes do Instagram. Para isso, foram analisadas as *hashtags* mais populares na rede social, através de ferramentas online, como por exemplo top-hashtags.com.

A análise de *hashtags* facilitou a escolha de um nicho e, por conveniência, foram escolhidos os nichos *Animals* (animais, especificamente *Dog*, cachorro), e *Food* (comida) para serem estudados paralelamente.

Tendo os nichos escolhidos, foi realizado uma **pesquisa por usuários** que postassem, em sua maioria, publicações pertencentes a *Dog* e *Food*. Através de recomendações encontradas em notícias, foram selecionados ao todo, 109 usuários do nicho *Dog* e 82 usuários do nicho *Food*.

Por se tratar de recomendações, não há garantia de que todos os usuários foram usuários populares (com um grande número de seguidores) na rede social, mas foi observado que o usuário com menor número de seguidores possuía 773 seguidores no nicho *Dog* e 6785 seguidores no nicho *Food*, enquanto o usuário com maior número de seguidores possuía 9595184 seguidores no nicho *Dog* e 247332 seguidores no nicho *Food*.

Tendo os usuários selecionados, um *crawler* em Python foi desenvolvido para se conectar com a LevPasha Instagram API, e então foram **coletados** dados relacionados aos usuários selecionados e suas publicações e armazenados em arquivos com a formatação JSON. Dentre os vários dados coletados, destaca-se: o número de seguidores de cada usuário, o número de curtidas de cada publicação e a imagem da publicação.

As publicações que possuíam seus dados coletados foram publicadas entre 01 de janeiro de 2015 até 30 de novembro de 2019 (data de realização da coleta), e continham apenas uma imagem por publicação, ou seja, vídeos e publicações

com mais de um item (imagem ou vídeo), foram descartados. Ao todo foram coletadas 179400 imagens do nicho *Dog* e 120353 imagens do nicho *Food*.

Foi percebido que muitas das imagens adquiridas não pertenciam aos nichos estudados, ou seja, existiam imagens com pessoas, textos, paisagens, dentre outros. Por isso, foi necessário o **pré-processamento** para limpar essas imagens do conjunto de dados.

Para isso, uma CNN com duas classes foi criada: para o nicho *Dog*, duas classes foram criadas, “dog” e “non-dog”, já para o nicho *Food*, as classes eram “food” e “non-food”. A ideia é que a CNN consiga diferenciar o que é *Dog* do que não é, e assim eliminar imagens que não fossem desse nicho (o mesmo para o nicho *Food*).

Porém, foi preciso o treinamento dessa CNN através de conjuntos de dados criados manualmente, e para facilitar esse processo foram utilizadas imagens aleatórias de *datasets* públicos da internet.

Mas, a possibilidade de imagens pertencentes às classes “non-dog” e “non-food” é infinita, por isso, foi pensado em escolher itens que poderiam aparecer no conjunto de dados do Instagram, sendo por exemplo: pessoas, objetos, paisagens, textos escritos, outros animais além de cachorros, dentre outros.

Para cada nicho, foram utilizadas 1900 imagens para cada classe em treino, e 700 imagens para cada classe em teste, totalizando 2600 imagens para cada classe no nicho *Dog* e outras 2600 imagens para cada classe no nicho *Food*. A escolha da quantidade de imagens é empírica.

Após a limpeza, o conjunto de dados do Instagram passou a ter 139016 imagens para o nicho *Dog* (reduzindo cerca de 22,5% do conjunto de dados) e 87732 imagens para o nicho *Food* (reduzindo cerca de 27,1% do conjunto de dados).

Na etapa de **processamento**, a abordagem utilizada se baseia em analisar imagens com um número de curtidas muito diferente, porém publicadas por usuários com um número de seguidores muito semelhante. Afinal, a tendência de duas imagens publicadas por usuários com número de seguidores semelhante, é de ter um número de curtidas também semelhante, logo há interesse em analisar essas duas imagens.

Mas, como determinar o quanto dois usuários possuem um número de seguidores muito semelhante, e o quanto duas imagens possuem o número de curtidas muito diferentes? Pensando nisso, um raio em porcentagem foi estipulado para cada comparação, considerando sempre o maior valor da comparação.

Para o caso do número de seguidores, o raio estipulado foi de 3%, e para o número de curtidas nas imagens, o raio estipulado foi de 95%. A escolha do valor para cada raio é empírica, visto que dependendo do raio escolhido, o número de comparações pode aumentar ou diminuir. Optou-se por utilizar porcentagem para que o raio se adeque a cada comparação, já que o número de seguidores e de curtidas é bem variado.

Por exemplo: comparando um usuário A com 98000 seguidores com um usuário B com 100000 seguidores, como B possui mais seguidores do que A, o raio de 3% é aplicado sobre B, resultando em 3000. Como o módulo da diferença entre

os seguidores de A e B é menor do que o valor obtido através do raio, considera-se que A e B são usuários com um número de seguidores muito semelhante.

Agora, comparando uma imagem I, pertencente a A, que possui 130000 curtidas e uma imagem J, pertencente a B, com 5000 curtidas, como I possui mais curtidas do que J, o raio de 95% é aplicado sobre I, resultando em 123500. Como o módulo da diferença entre as curtidas de I e J é maior do que o valor obtido através do raio, considera-se que I e J são imagens com um número de curtidas muito diferentes.

As comparações foram armazenadas no MongoDB. Para o nicho *Dog*, 1869770 comparações foram obtidas, e para o nicho *Food*, 1128893 comparações foram obtidas. Para cada comparação armazenada, a imagem com maior número de curtidas foi considerada interessante, enquanto a imagem com menor número de curtidas foi considerada desinteressante.

Mas, a mesma imagem pode aparecer em diversas comparações, assim como pode ser considerada interessante em uma comparação e desinteressante em outra. Para que haja um rótulo fixo para cada imagem, foi contado o quanto cada imagem é considerada interessante e desinteressante em todas as comparações que apareceram, e assim seu rótulo foi definido majoritariamente. Em caso de empate, um rótulo aleatório foi definido.

Porém, em ambos os nichos a quantidade de comparações foi muito grande, e contabilizar todas as linhas para obter todas as imagens exigiria uma grande demanda de tempo e processamento. Por essa razão, cerca de 50% das comparações tiveram suas imagens contabilizadas, o que resultou em 8485 imagens interessantes e 3478 desinteressantes para o nicho *Dog*, e 3384 imagens interessantes e 2617 desinteressantes para o nicho *Food*. Essa quantidade foi julgada suficiente para o treinamento da CNN.

O resultado do processamento gerou um conjunto de dados de imagens rotuladas como interessantes e desinteressantes para cada nicho. Esses conjuntos de imagens foram utilizados no treinamento de uma CNN para que ela automatize a classificação de novas imagens.

A CNN utilizada na etapa de **classificação** é semelhante à utilizada na etapa de pré-processamento, utilizando a arquitetura VGG16 unido a um modelo de classificação de regressão logística. Mas, para que todo o conjunto de dados seja melhor aproveitado, foi realizada a validação cruzada com o método *k-fold*, em que o conjunto de dados foi dividido entre 10 subconjuntos, sendo cada subconjunto utilizado uma vez como teste, enquanto os outros 9 restantes utilizados como treino do algoritmo. Dessa forma, foram realizados ao todo 10 testes.

RESULTADOS E DISCUSSÃO

As Figuras 3 e 4 mostram exemplos de imagens que foram consideradas interessantes e desinteressantes para o treinamento da CNN no nicho *Dog*.

Figura 3 – Imagens consideradas interessantes para o nicho *Dog*



Fonte: Autoria Própria (2020).

Figura 4 – Imagens consideradas desinteressantes para o nicho Dog



Fonte: Autoria Própria (2020).

Dessa forma, pode-se observar que há um padrão para imagens interessantes e desinteressantes nesse nicho, de modo que as imagens interessantes, em sua maioria, apresentam fotos mais profissionais de cachorros, em posições que podem encantar os usuários e até mesmo com cachorros de raças, enquanto as imagens consideradas desinteressantes, apresentam montagens e são muito poluídas visualmente.

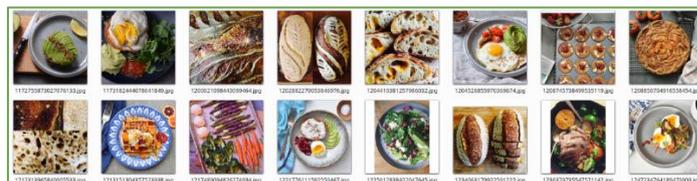
Já as Figuras 5 e 6 mostram exemplos de imagens que foram consideradas interessantes e desinteressantes para o treinamento da CNN no nicho Food.

Figura 5 – Imagens consideradas interessantes para o nicho Food



Fonte: Autoria Própria (2020).

Figura 6 – Imagens consideradas desinteressantes para o nicho Food



Fonte: Autoria Própria (2020).

Neste caso, o padrão acaba sendo mais difícil de ser reconhecido, mas pode-se dizer que as imagens consideradas interessantes apresentam uma estética melhor na apresentação das comidas, o que atrai o paladar humano, enquanto as imagens consideradas desinteressantes acabam não sendo tão apetitosas visualmente.

As métricas de cada execução da CNN da etapa final de classificação são apresentadas na Tabela 1.

Tabela 1 – Métricas da CNN da etapa de classificação

Nicho	Classe	Acurácia (%)	Precisão (%)	Recall (%)	F1 Score (%)
<i>Dog</i>	interessante	95	92	95	93
	desinteressante	80	86	80	83
<i>Food</i>	interessante	85	83	85	84
	desinteressante	77	80	77	79

Fonte: Autoria Própria (2020).

A acurácia total do nicho *Dog* foi de 91%, e a acurácia total do nicho *Food* foi de 82%. Comparando com a CNN de (ROSEBROK, 2019) que classifica comidas e não comidas e que obteve uma acurácia de 98%, pode-se dizer que o resultado é satisfatório, pois por mais que o objetivo da classificação seja diferente, o modelo utilizado é semelhante, assim, os resultados obtidos foram próximos, considerando que é mais difícil classificar o que é interessante e desinteressante do que é comida e não comida.

Observa-se que em ambos os nichos houve um acerto maior em imagens interessantes se comparadas com imagens desinteressantes, porém, observa-se que o nicho *Dog* obteve uma acurácia maior que o nicho *Food*. Isso pode ter ocorrido por seu conjunto de imagens ser mais específico, tornando mais fácil para o algoritmo determinar *dog* e *non-dog* do que *food* e *non-food*.

CONCLUSÕES

O objetivo do trabalho foi criar uma CNN capaz de aprender e classificar imagens interessantes e desinteressantes como recomendações para a rede social Instagram. Para a análise e validação da proposta, foram elencados os nichos *Dog* e *Food*. No entanto, vale ressaltar que o método proposto é passível de generalização a quaisquer outros nichos. Tendo isso em mente, pode-se dizer que esta é uma pesquisa inicial e que por isso pode-se dizer que seu resultado foi satisfatório.

Há melhorias que podem ser realizadas em trabalhos futuros, como encontrar técnicas para identificar os melhores usuários em um determinado nicho, melhorar a base de dados da etapa de pré-processamento para que a limpeza seja mais efetiva, utilizar outro tipo de análise no momento da seleção de publicações interessantes e desinteressantes, assim como utilizar nessa análise outras informações provindas do conjunto de dados e, por fim, utilizar outros modelos de CNNs, outros métodos de transferência de aprendizado e outros modelos de aprendizado de máquina para a classificação.

AGRADECIMENTOS

Meus agradecimentos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), à Fundação Araucária e a Universidade Tecnológica Federal do Paraná (UTFPR-CP), pela oportunidade de realizar esse projeto de pesquisa.

REFERÊNCIAS

BUYA, R.; CALHEIROS, R.; DASTJERDI, A. **Big Data: Principles and Paradigms**. Elsevier Science, 2016. ISBN 9780128093467. Disponível em: <https://books.google.com.br/books?id=MfOeCwAAQBAJ>. Acesso em: 22 jul 2020.

MELO, E. **Sistema de Recomendação de Imagens Baseado em Atenção Visual**. Tese de doutorado do curso de Ciência da Computação - Universidade Federal de Uberlândia, Uberlândia, 2016. Disponível em: <https://repositorio.ufu.br/bitstream/123456789/17824/1/SistemaRecomendacaoImagens.pdf>. Acesso em: 22 jul 2020.

MENDES, D. **Big Data Fundamentos 2.0**. Data Science Academy, Brasília, 2016. Disponível em: <https://www.datascienceacademy.com.br/course?courseid=bigdata-fundamentos>. Acesso em: 22 jul 2020.

PRABHU, R. **Understanding of Convolutional Neural Network (CNN) — Deep Learning**. 2018. Disponível em: <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>. Acesso em: 22 jul 2020.

ROCHA, E. **O Que Significa Engajamento?**. Ignição Digital, Brasília, 2019. Disponível em: <https://www.ignicaodigital.com.br/o-que-significa-engajamento/>. Acesso em: 22 jul 2020.

ROSEBROK, A. **Transfer Learning with Keras and Deep Learning**. PYIMAGESEARCH, 2019. Disponível em: <https://www.pyimagesearch.com/2019/05/20/transfer-learning-with-keras-and-deep-learning/>. Acesso em: 22 jul 2020.

SANTANA, M. **Deep Learning para Sistemas de Recomendação (Parte 1) — Introdução**. Data Hackers, 2018. Disponível em: <https://medium.com/data-hackers/deep-learning-para-sistemas-de-recomenda%C3%A7%C3%A3o-parte-1-introdu%C3%A7%C3%A3o-b19a896c471e>. Acesso em: 22 jul 2020.

SBCOACHING. **Engajamento: o que é, importância e como criar**. SBCoaching, São Paulo, 2019. Disponível em: <https://www.sbcoaching.com.br/blog/engajamento>. Acesso em: 22 jul 2020.

TAKAHASHI, M. **Estudo comparativo de Algoritmos de Recomendação**. Monografia do curso de Bacharelado em Ciências da Computação - Universidade de São Paulo, São Paulo, 2015. Disponível em: https://bcc.ime.usp.br/tccs/2014/marcost/monografia_final.pdf. Acesso em: 22 jul 2020.