Projeto de Iniciação Científica

Edital 04/2022

Título do projeto: SincNet para Classificação de Instrumentos Musicais

Palavras-Chave: Processamento Digital de Sinais, Processamento Digital de Áudio, Aprendizado Profundo, Classificação de Padrões.

RESUMO

Problemas de classificação constituem uma parte importante do campo de processamento digital de sinais. A capacidade de desenvolver sistemas que reconheçam padrões em sinais e sejam capazes de os identificar tem sido aproveitada em inúmeras situações, como por exemplo: reconhecimento biométrico, reconhecimento de voz e faces e até mesmo escrita. Diversos tipos de abordagens podem ser adotadas para solucionar esses problemas, combinando diferentes tipos de algoritmos de aprendizado de máquina (ML, *Machine Learning*). Entre eles, uma abordagem muito utilizada é através do aprendizado profundo (DL, *Deep Learning*), que tem como objetivo estruturar os algoritmos de forma mais complexa e robusta utilizando redes neurais artificiais . Entretanto, essa complexidade pode resultar em um elevado custo computacional na execução dos algoritmos, também ocasionando *overfitting* em alguns casos. Sendo assim, o presente trabalho tem como objetivo realizar a classificação de instrumentos musicais utilizando redes neurais convolucionais (CNN, *Convolutional Neural Networks*), explorando uma estrutura modificada, denominada SincNet, que utiliza filtros sinc na primeira camada da CNN, como forma de reduzir o número de parâmetros e complexidade da rede - realizando a avaliação de comparação de desempenho deste método em relação à modelos de CNNs tradicionais.

1 INTRODUÇÃO

Com o desenvolvimento de diferentes formas de compartilhamento de informação, o campo de estudo de processamento de sinais apresentou um grande crescimento nos últimos tempos. Diversas formas de gerar e transmitir dados foram criados e popularizados, de forma que esse fluxo de informação gradativamente tornou-se parte do cotidiano da sociedade. Enviar e receber sinais de áudio e vídeo é indiscutivelmente parte da vida de grande parcela da população, logo estudar métodos eficazes de realizar o tratamento e transmissão desses sinais é uma tarefa valiosa para a otimização desse processo. É neste ponto que a área de processamento de sinais atua - estudando e desenvolvendo maneiras para tratar de diferentes tipos de sinais.

Nesse contexto, o aprendizado profundo (DL, *Deep Learning* [1], uma vertente da área de ML, tem sido cada vez mais explorado para se obter soluções para o processamento de sinais. Um exemplo são as Redes Neurais Convolucionais (CNN, *Convolutional Neural Networks*, um tipo de rede neural diretamente associada à noção de DL originalmente proposta para o processamento de imagens [2]. Na figura 1, pode ser vista uma ilustração da estrutura básica de uma CNN - composta por diferentes camadas de processamento. Essencialmente, o conjunto de camadas iniciais realiza operações de **convolução** com filtros que permitem extrair características específicas da imagem que auxiliarão as camadas próximas à saída da rede a realizar a classificação correta do sinal de entrada. Em muitos casos, essas camadas próximas à saída da rede, estão estruturadas de forma análoga a uma rede neural MLP (**multilayer perceptron**) tradicional.

Embora tenham sido tradicionalmente propostas para o processamento de imagens, as CNNs também foram adaptadas para o processamento de outros tipos de dados, e assim, algoritmos que envolvem CNN têm sido cada vez mais utilizados como forma de realizar tarefas na área de processamento de sinais, obtendo resultados expressivos em diversos casos. Diferentes estudos explorando CNNs para o processamento de sinais de áudio têm revelado o grande potencial das redes em extrair características relevantes diretamente das amostras do sinal [3] [4]. Nesse contexto, o presente trabalho tem como objetivo investigar o uso das CNNs na solução do pro-

blema de classificação de instrumentos musicais.

A investigação a ser realizada tem como base uma trabalho recente [5] no qual é proposta uma alteração na estrutura da CNN, explorando ideias bem conhecidas em técnicas de processamento de sinais *clássicas*, a fim de tornar a rede mais enxuta e mais especializada para aplicações em sinais de áudio. A nova estrutura, denominada de *SincNet* - também demonstrada na Figura 1, processa os dados de maneira diferente na primeira camada da CNN. Neste caso, em vez de definir um grande conjunto de filtros diferentes para o processo de convolução, a rede utiliza apenas um tipo de filtro padrão baseado na função $sinc$ - uma função de grande importância na área de processamento digital de sinais, muito utilizada no processo de reconstrução de sinais, dada por $\frac{sen(x)}{x}$. Dessa forma, os filtros utilizados na primeira cada da CNN são definidos por meio de apenas dois parâmetros, reduzindo drasticamente a quantidade de parâmetros a serem ajustados na rede. Isso permite que o algoritmo adapte adequadamente a rede mesmo operando com uma quantidade reduzida de dados de treinamento, além de permitir, mais facilmente, interpretar qual a representação dos sinais é explorada pela rede.

A grande flexibilidade das CNNs, entretanto, está associada um grande número de parâmetros a serem ajustados, o que acaba exigindo um grande volume de dados para que o treinamento seja feito adequadamente em alguma aplicações, como no processamento de sinais unidimensionais (como no caso de sinais de áudio). Embora se destaquem justamente pela sua alta complexidade, esses algoritmos precisam selecionar uma grande quantidade de filtros para realizar o processo de convolução, o que exige que esses filtros sejam definidos previamente, de forma que seja necessário definir uma grande quantidade de parâmetros para a computação de resultados. Assim, ao lidar com a classificação de sinais de áudio, utilizar os algoritmos clássicos de CNN pode não ser a solução mais eficiente.

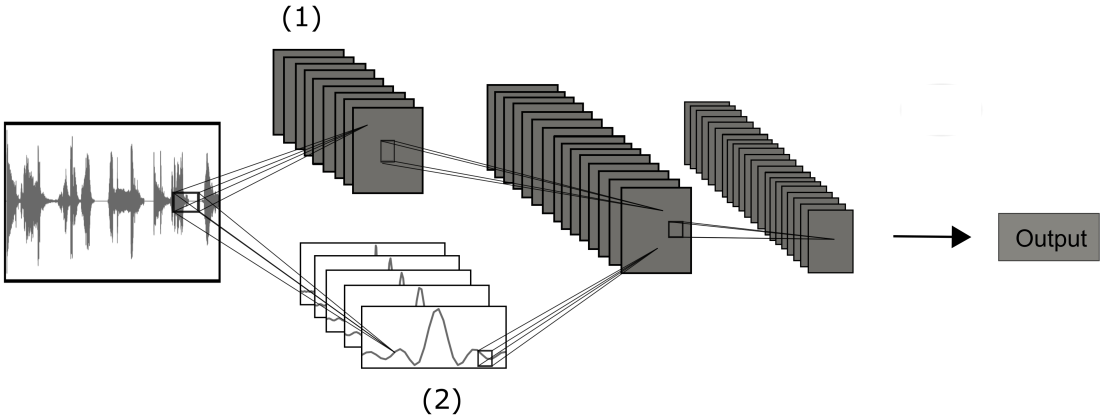

Figura 1: Esquemático do funcionamento de uma CNN tradicional em comparação com uma rede utilizando filtros SincNet. Em (1), temos que a primeira camada é composta por diferentes *kernels*, não possuindo nenhuma especificação previamente definida, conforme ocorre em uma CNN clássica. Já em (2), a primeira camada é composta por uma série de filtros sinc em frequências variadas, de forma que os únicos parâmetros a serem definidos serão as faixas de frequências admitidas para o caso, conforme descreve o modelo SincNet.

A SincNet foi desenvolvida para atuar no problema de classificação de sinais de voz, entretanto diferentes trabalhos já adaptaram o algoritmo para que funcione em diferentes problemas, como por exemplo no processamento de sinais de eletroencefalograma (EEG) [6] e até mesmo para a detecção de falhas em motor de indução [7]. No entanto, ao se tratar da classificação de instrumentos musicais, ainda são utilizados métodos que possuem grande dependência em características extraídas por métodos mais tradicionais de processamento de sinais, como os coeficientes mel-cepstrais (MFCC, *Mel Frequency Cepstral Coefficients*) [8].

2 OBJETIVOS

O principal objetivo deste trabalho é realizar a implementação de um algoritmo utilizando a SincNet para realizar a classificação de instrumentos musicais, realizando a comparação de desempenho com o a rede convolucional tradicional.

Como objetivos específicos, temos:

- Estudo dirigido na área de aprendizado de máquina, em particular em estruturas associadas ao aprendizado profundo;

- Estudo dirigido na área de processamento digital de sinais, com foco na definição de filtros digitais;
- Desenvolver habilidades de implementação de algoritmos de aprendizado profundo em Python, explorando bibliotecas como Keras e Tensorflow;
- Construir e/ou organizar uma base de dados para treinamento e teste do sistema de classificação de instrumentos musicais;
- Desenvolver sistema de classificação baseado na rede sincnet, e comparar o desempenho com abordagens mais tradicionais;

3 METODOLOGIA

Tendo em vista a relação direta entre o ML e o DL, primeiramente, será realizado um estudo a respeito dos principais tópicos de ML, como forma de compreender os princípios dos algoritmos de aprendizado de máquina. Além disso, também serão estudados implementações deste algoritmos em Python, bem como as principais bibliotecas utilizadas no estudo da área, como por exemplo: TensorFlow, scikit-learn e Keras. Em seguida, será realizado um aprofundamento em tópicos de de DL, que é base fundamental para a compreensão dos conceitos de CNN.

Além do estudo na área de inteligência artificial, também será necessário compreender a teoria de filtros digitais, fundamental para a compreensão da SincNet. Para isso, também será necessário realizar estudos em processamento digital de sinais, entendendo desde a amostragem de sinais, transformadas e por fim, com a ênfase principal na compreensão dos filtros digitais. Com o estudo de ambas as áreas, torna-se possível analisar o método de classificação aplicando a SincNet. Através da revisão bibliográfica, também pode ser possível determinar possíveis diferenças e melhorias na implementação das redes que podem ser incorporadas e testadas neste trabalho.

Para o sistema de classificação, serão selecionados base de dados específicas para a execução de testes de classificação de instrumentos musicais, sendo que além do algoritmo de classificação utilizando a SincNet, também serão realizadas aplicações de algoritmos de classificação que utilizem CNNs tradicionais, como forma de obter um resultado comparativo. A avaliação será feita com base no erro que os

algoritmos tiverem nas tentativas de classificação. Todos os testes e implementação irão utilizar as bibliotecas citadas anteriormente - em Python.

4 EXEQUIBILIDADE

O presente trabalho é baseado em estudos teóricos voltado para as áreas de DL e processamento digital de sinais, bem como a implementação computacional destes algoritmos para a realização dos teste de classificação. Como a implementação será realizada em Python, todas as ferramentas computacionais necessárias serão de código aberto, logo não será necessário utilizar *softwares* comerciais.

Serão utilizadas bases de dados gratuitas para realizar o processo de classificação, de forma que não seja necessário realizar esforços para a coleta de dados para os testes. Assim, o trabalho pode ser realizado sem qualquer tipo de equipamento, apenas com o uso de computadores domésticos, sem a necessidade de uso qualquer espaço físico específico.

5 CRONOGRAMA

Para a realização do projeto, o trabalho foi estruturado por partes, nas seguintes etapas:

1. Revisão Bibliográfica: revisão de estudos realizados em áreas relacionadas, como classificação utilizando SincNet e classificação de instrumentos musicais;
2. Estudo sobre Processamento Digital de Sinais;
3. Estudo sobre Aprendizado Profundo;
4. Estudo sobre linguagem Python: estudar as bibliotecas e recursos necessários para a implementação do
5. Preparação de testes de classificação: separação do banco de dados e estruturar a fase de testes;
6. Implementação e avaliação do algoritmo de classificação por SincNet;
7. Redação do Relatório;

Atividades	Mês											
	1	2	3	4	5	6	7	8	9	10	11	12
Revisão Bibliográfica	■	■										
Estudo sobre PDS			■	■	■	■	■					
Estudo sobre Aprendizado Profundo			■	■	■	■	■					
Estudo sobre linguagem Python			■	■								
Preparação de testes de classificação						■	■					
Implementação e execução de algoritmo								■	■	■		
Redação do Relatório											■	■

Referências

[1] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. http://www.deeplearningbook.org.

[2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

[3] S. Hershey, S. Chaudhuri, D. P. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, M. Slaney, R. J. Weiss, and K. Wilson, "CNN architectures for large-scale audio classification," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 131–135, Institute of Electrical and Electronics Engineers Inc., jun 2017.

[4] H. Lee, L. Yan, P. Pham, and A. Y. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Advances in Neural Information Processing Systems 22 - Proc. 2009 Conference*, pp. 1096–1104, 2009.

[5] M. Ravanelli and Y. Bengio, "Speaker Recognition from Raw Waveform with SincNet," in *2018 IEEE Spoken Language Technology Workshop, SLT 2018 - Proceedings*, pp. 1021–1028, IEEE, feb 2019.

[6] H. Zeng, Z. Wu, J. Zhang, C. Yang, H. Zhang, G. Dai, and W. Kong, "EEG emotion classification using an improved sincnet-based deep learning model," *Brain Sciences*, vol. 9, nov 2019.

[7] F. B. Abid, M. Sallem, and A. Braham, "Robust interpretable deep learning for intelligent fault diagnosis of induction motors," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 3506–3515, 2020.

[8] D. G. Bhalke, C. B. Rao, and D. S. Bormane, "Automatic musical instrument classification using fractional fourier transform based- MFCC features and counter propagation neural network," *Journal of Intelligent Information Systems*, vol. 46, pp. 425–446, jun 2016.

[9] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, May 2015.