**Detector de motores a combustão por meio de análise sonora através de uma rede neural.**

Matheus Felipe Sozza¹\*;Maurício Eloy2

1 Robert Bosch GmbH. Engenheiro Eletricista. Rodovia Anhanguera, km 98 – Vila Boa Vista; 13065-900 Campinas, São Paulo, Brasil

2 Nome da Empresa ou Instituição (opcional). Titulação ou função ou departamento. Endereço completo (pessoal ou profissional) – Bairro; 00000-000 Cidade, Estado, País

\*autor correspondente: matheussozza@hotmail.com

**Detector de motores a combustão por meio de análise sonora através de uma rede neural.**

**Resumo (ou Sumário Executivo)**

O resumo é uma descrição geral do trabalho, apresentando de forma sucinta todas as seções do TCC. **No caso do curso de** **MBA em Gestão de Projetos**, em função das particularidades do trabalho – que pode ser um plano de projeto – **é facultado aos alunos o uso do termo** **Sumário Executivo**. Essa seção deve ser escrita de forma clara e objetiva, informando ao leitor em que consiste o TCC e despertando o seu interesse para a leitura de todo o trabalho. Os termos Resumo (ou Sumário Executivo) e Palavras-chave devem ser grafados em negrito, alinhados à esquerda com apenas a primeira letra da palavra em letra maiúscula. O Resumo ou Sumário Executivo deve conter no máximo 250 palavras, sendo redigido em um único parágrafo, em espaçamento simples e no tempo verbal pretérito perfeito do indicativo (passado). O início da seção deve conter uma curta apresentação sobre a importância/justificativa (contextualização do tema) do trabalho. Em seguida, deve-se apresentar o objetivo geral, que deve ser redigido de maneira sucinta e direta. Ademais, a seção deve conter uma breve descrição da metodologia empregada na pesquisa, abordando os aspectos mais importantes para o entendimento do TCC. É importante destacar os principais resultados obtidos a partir da pesquisa, apresentando relações e/ou considerações que chamem a atenção do avaliador e do público em geral. O final do resumo (ou sumário executivo) deve conter uma conclusão geral (com uma ou duas frases). Deve-se ter cuidado para não repetir a descrição dos resultados nesta seção.

**Palavras-chave:** (inserir até cinco palavras diferentes das contidas no título, separadas por ponto-e-vírgula).

**Título em inglês** **ou espanhol** (opcional)

**Abstract** ou **Resumen** (opcional)

....................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................

**Keywords** ou **Palabras Clave:** (opcional)

**Introdução**

Motores a combustão são sistemas muito presentes no cotidiano dado sua extensa aplicabilidade e o fato de ser uma tecnologia já há muito conhecida e cujo fenômeno base, a combustão, é dominada com robustez pela indústria.

Ao mesmo tempo, é sabido que apesar da relevância que esses dispositivos possuem, existem alguns pontos contra notáveis, como a poluição atmosférica e consequente potencialização do efeito estufa e aquecimento global, além dos riscos à saúde das pessoas expostas direta e prolongadamente aos poluentes emitidos pelo escapamento (Nikischer, 2020).

Fazendo uso dos padrões sonoros do motor é possível extrair informações a respeito da sua performance de operação, o que permite julgar se o dispositivo está em funcionamento, operando corretamente, e até mesmo diagnosticar falhas em partes específicas, como citado em Wu. Z. et al., (2022). Atualmente o diagnóstico de falhas é subjetivo e altamente correlacionado à expertise do técnico mecânico responsável pela avaliação e manutenção do veículo, o que resulta em baixo nível de sucesso, havendo então uma oportunidade de melhoria desse processo por meio de algoritmos inovadores (Kemalkar e Bairagi, 2016).

Outra possível utilização de dados sonoros de motores é aquela voltada à geração de estatísticas de tráfego e planejamento de demanda, ou até mesmo no controle de semáforos baseado em situações especiais, como a presença de veículos de emergência (ambulâncias), como destacado em Analytics Vidhya (2022). Dados extraídos num contexto de Big-Data têm sido cada vez mais utilizados no âmbito das cidades inteligentes (*Smart-Cities*), auxiliando no planejamento e gestão do tráfego das zonas rural e urbana (Zhao, Y. et al., 2018).

Mohammadi e Al-Fuqaha (2018) enunciam que, apesar das altas capacidades de coleta e armazenamento de dados dos sistemas computacionais atuais, muito pouco se aproveita da informação ali contida, com métodos tradicionais baseados apenas em análise temporal, os quais negligenciam a presença de padrões valorosos contidos nos dados armazenados e que não visíveis diretamente numa análise por amostras. Por outro lado, o uso de redes neurais profundas (DNN ou *Deep Neural Networks*) se mostra uma alternativa válida e promissora para a extração de informações perspicazes do ponto de vista analítico.

Outro ponto enunciado por Mohammadi e Al-Fuqaha (2018) é o fato de que, uma vez que os dados são coletados e armazenados, é improvável que venham a ser reutilizados no futuro, encorajando-se o processamento imediato.

O presente trabalho visa demonstrar como é possível extrair valor de amostras sonoras por meio do processamento intensivo dos dados através redes neurais convolucionais [CNNs], mesmo quando os padrões e métricas a se extrair não são observáveis ou perceptíveis a priori para um processamento através de algoritmos tradicionais e explicáveis, os quais se baseiam em regras pré-estabelecidas e modelagem baseada puramente em premissas matemáticas. (Bhatia, 2018).

**Material e Métodos**

**Dados**

Para o trabalho em questão foi utilizada a base de dados ‘UrbanSound8k’. Essa coletânea consiste em 8732 recortes sonoros de até 4 segundos de duração os quais já estão previamente rotulados em 10 categorias distintas, sendo elas: Ar-Condicionado, Buzina de Carro, Crianças brincando, Cães latindo, Perfuração, Motor em marcha lenta, Tiro de arma de fogo, Britadeira, Sirene e Música Urbana.

Todos os recortes de áudio são provenientes de amostras disponíveis no website [www.freesound.org](http://www.freesound.org), sendo distribuídas em 10 diferentes pastas de maneira a facilitar a divisão dos dados para treino e teste quando aplicados em algoritmos de Machine-Learning [ML] e/ou Deep-Learning [DL]. É importante manter a distribuição original das 10 pastas, evitando a mistura e redivisão das amostras, pois dessa maneira evita-se que subamostras originárias de um mesmo arquivo de áudio sejam utilizadas tanto para o treino como para o teste e validação do modelo, consequentemente gerando resultados artificialmente altos, como recomendado em Salamon et. al. (2014) para garantir a validades dos resultados. Os dados estão balanceados dentro de cada uma das dez pastas, isto é, todas as dez categorias rotuladas (Ar-Condicionado, Buzina, Crianças, etc.) estão presentes de maneira bem distribuída.

Os recortes de áudio foram fornecidos em formato ‘.wav’ e com taxa de amostragem de 44100 Hertz.

**Pré-processamento – Segmentação e balanceamento de classes**

De maneira a normalizar e preparar os recortes sonoros para o posterior processamento no algoritmo de ML, é importante adotar algumas premissas que serão mantidas durante todo o processo no que se refere à segmentação e amostragem dos recortes sonoros originais.

A duração da amostra padrão foi arbitrada em 500 milissegundos ou 1 segundo (a depender do experimento executado). As amostras provenientes da base de dados ‘Urbansound8k’ cuja duração é menor que a duração padrão foram descartadas.

Dado o escopo do trabalho, a classificação dos dados se dá de maneira binária, isso é, o intuito é classificar apenas se há um motor a combustão em funcionamento ou não no recorte sonoro em questão, não sendo necessária a classificação em dez diferentes rótulos como originalmente presente na base de dados “UrbanSound8k”.

Para tal, as dez categorias originais da base de dados foram reduzidas em apenas duas. A primeira categoria foi chamada de Motores a combustão [MC], onde foi considerado o rótulo original “Motor em Marcha Lenta” apenas; A segunda categoria foi chamada “Não-Motor a Combustão” [NMC], onde todos os demais rótulos foram agrupados.

Por consequência, gerou-se um notável desbalanceamento dos dados, onde 80 a 90% das amostras estiveram contidas na categoria NMC, e apenas 10 a 20% das amostras estiveram contidas na categoria MC. No que tange os algoritmos de ML e DL, assume-se que os dados das classes consideradas estão distribuídos razoavelmente de maneira similar e balanceada, o que muitas vezes não é verdade num contexto de vida real. Tem-se então um fator dificultador ao aprendizado do algoritmo que tende a apresentar um elevado viés em direção ao grupo majoritário, quando o interesse é realizar previsões de qualidade relativas à classe minoritária (Krawczyk, 2016).

Para atingir o objetivo de balancear as classes de maneira equilibrada, direcionou-se o foco desta etapa de pré-processamento à classe minoritária MC e aplicou-se uma série de técnicas de maneira a sobreamostrar e criar artificialmente dados dentro da mesma, técnicas popularmente conhecidas como “oversampling” [OS] e “data-augmentation” [DA].

Durante a segmentação em amostras padrão da classe MC, aplicou-se um fator de “overlapping” de maneira a aumentar artificialmente o número de amostras contidas na classe em questão, como referenciado em Chen (2019) e Song (2021). Neste processo, realiza-se não apenas a segmentação da amostra original, mas existe uma sobreposição entre uma determinada subamostra e as respectivas subamostras antecessora e sucessora, de maneira que parte do conteúdo é compartilhado entre as mesmas. O processo é mais bem ilustrado na Figura 1.

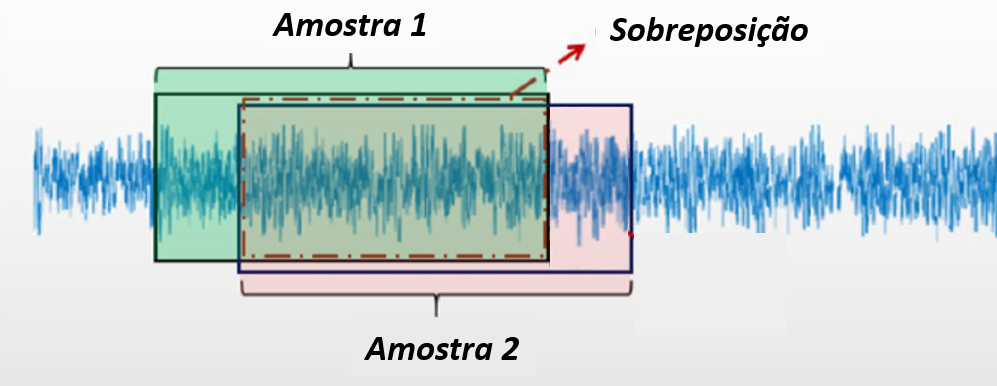


Figura . Demonstração gráfica do processo de sobreamostragem ou "oversampling" com sobreposição ou "overlapping"  
Fonte: Song (2021) – traduzido e adaptado

Para uma dada amostra de áudio, é possível calcular o número de subamostras resultantes após a segmentação através da eq. (1):

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

onde, T: é a duração da amostra original a ser segmentada; Ta: é a duração alvo das subamostras; S: é o fator de sobreposição que pode variar entre 0 (sem sobreposição) e 1 (sobreposição total); e N: é o número de subamostras resultantes, arredondado sempre para baixo.

Assim sendo, arbitrando-se o tamanho das subamostras e o fator de sobreposição, é possível prever o quanto o conjunto de dados aumentará com base na técnica de segmentação com sobreposição.

**Pré-processamento – Data Augmentation**

Uma vez que os dados resultantes da segmentação com sobreposição contêm informações idênticas e repetidas (pois é uma técnica de sobreamostragem), aumenta-se então a tendência ao sobreajuste ou “overfitting” do nosso modelo devido ao aprendizado reforçado de um mesmo tipo de padrão repetidamente, conforme mencionado em Fernandez (2018). Essa situação é caracterizada por uma alta performance de classificação obtida nos dados de treino em oposição à uma baixa performance obtida durante a etapa de teste do modelo, o que indica que o algoritmo é muito pouco confiável ao ser aplicado em dados nunca vistos.

Para evitar tal problema, duas técnicas de “Data Augmentation” são aplicadas, com o intuito de criar subamostras sintéticas baseadas nas amostras previamente existentes: Compressão ou Dilatação no Tempo e Adição de Ruído Branco Gaussiano.

A Compressão ou Dilatação no Tempo é realizada de maneira simples e direta, onde uma determinada amostra de áudio é comprimida ou dilatada, de maneira a ser executada de maneira mais rápida ou mais lenta, a exemplo do que é realizado em Zhou (2022). Dado que esse processo altera a duração da amostra, a aplicação do mesmo é realizada em coordenação com a etapa de segmentação e sobreposição, de maneira a manter inalterada a duração das subamostras resultantes, estando de acordo com a duração padrão previamente arbitrada. Por exemplo, um áudio de 30 segundos dilatado com um fator de 2 teria uma nova duração de 60 segundos; do contrário, um áudio de 30 segundos comprimido com um fator de 0.5 teria uma nova duração de 15 segundos (Zhou, 2022).

Já a Adição de Ruído Branco Gaussiano é realizada por meio da adição à amostra original de um sinal aleatório contendo intensidade uniforme em diferentes frequências, possuindo uma densidade espectral de potência constante. Uma vez que a mesma amostra dá origem a diferentes subamostras artificiais com diferentes níveis de ruído aleatório, espera-se que o modelo seja mais robusto ao se deparar com tal tipo de ruído em um futuro dado de entrada, o que melhora sua capacidade de generalização e performance como um todo, evitando um sobreajuste aos dados de treino. Como citado por Bishop (1995), em algumas circunstâncias é esperada uma melhora significativa na performance de generalização, pois treinar o modelo com ruído é equivalente à uma forma de regularização dos dados na qual um termo extra é adicionado à função de erro.

**Resultados e Discussão**

O título da seção Resultados e Discussão deve ser alinhado à esquerda, grafado em negrito com as primeiras letras das palavras em letras maiúsculas. É permitido que a seção seja dividida em subtópicos com formatação de acordo com a descrição no item 1.1 Formato e margens, apresentados na mesma ordem da seção Material e Métodos. Nesta seção devem ser apresentados, discutidos e interpretados os resultados obtidos no trabalho, ou seja, autores devem fazer uma discussão comparativa dos resultados do seu trabalho com aqueles existentes na literatura e elaborar uma análise crítica dos dados, destacando as limitações e pontos positivos dos resultados.

**Conclusão(ões) ou Considerações Finais**

O título da seção Conclusão(ões) ou Considerações Finais deve ser alinhado à esquerda e grafado em negrito. Fica a critério do aluno e do orientador a escolha de qual termo melhor se adequa ao trabalho. Esta seção deve conter frases curtas, apresentando as conclusões e inferências elaboradas a partir da discussão dos resultados. É importante que estas frases não sejam meras reproduções dos resultados, respondendo aos objetivos propostos no trabalho. Os autores não devem, em hipótese alguma, mencionar, citar ou reproduzir resultados de outros estudos na(s) conclusão(ões) ou considerações finais do TCC. Por fim, salienta-se que essa seção não deve conter tabelas ou figuras, sendo redigida de forma sucinta.

**Agradecimento** (opcional, 1 parágrafo, bem sucinto)

O título da seção Agradecimentos deve ser alinhado à esquerda e grafado em negrito, primeira letra da palavra grafada em letra maiúscula. Trata-se de seção opcional, de no máximo três linhas, na qual o autor agradece aqueles que contribuíram de maneira relevante para o desenvolvimento do trabalho e elaboração do TCC, mas que não tiveram o envolvimento intelectual necessário à atribuição de coautoria do mesmo, abstendo-se totalmente da menção ou citação de nomes de empresas, instituições ou pessoas que permitiram ou contribuíram com o desenvolvimento do trabalho, a menos que esteja documentalmente autorizado a fazê-lo.

**Referências**

Neste tópico deverá conter todas as referências dos trabalhos citados no texto e formatadas seguindo rigorosamente as normas do MBA USP ESALQ. Para mais informações, vide o manual de “Normas para Elaboração do Trabalho de Conclusão de Curso” disponível no sistema TCC.

Analytics Vidhya. 2022. Vehicle Sound Classification Using Deep Learning. Disponível em <https://www.analyticsvidhya.com/blog/2022/01/vehicle-sound-classification-using-deep-learning>. Acesso em 3 de outubro de 2022.

Bishop, C. 1995. Training with Noise is Equivalent to Tikhonov Regularization. Neural Computation, vol.7, no.1: 108-116.

Bhatia, R. 2018. How do Machine Learning algorithms differ from traditional algorithms? Analytics India Magazine. Disponível em: https://analyticsindiamag.com/how-do-machine-learning-algorithms-differ-from-traditional-algorithms/. Acesso em 22 dez. 2022.

Chen, H; Hu, N; Cheng, Z. 2019. A deep convolutional neural network based fusion method of two-direction vibration signal data for health state identification of planetary gearboxes. Measurement, Volume 146. 268-278.

Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., Herrera, F. 2018. Learning from imbalanced data sets. Springer, Berlin, Alemanha.

Kemalkar, A.K.; Bairagi, V.k.. 2016. Engine Fault Diagnosis Using

Sound Analysis. International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT): 943-946

Krawczyk, Bartosz. 2016. Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence 5 (4): 221-232.

Mohammadi, M.; Al-Fuqaha, A. 2018. Enabling Cognitive Smart Cities Using Big Data and Machine Learning: Approaches and Challenges. IEEE Communications Magazine, vol. 56, no. 2: 94-101.

Nickischer, A. 2020. Environmental Impacts of Internal Combustion Engines and

Electric Battery Vehicles. D.U. Quark, Volume #4 (Issue #2): 21-31.

Salamon, J; Jacoby, C; Bello, J. 2014. A Dataset and Taxonomy for Urban Sound Research. 22nd ACM International Conference on Multimedia, 2014, Orlando, Flórida, Estados Unidos.

Song, X; Cong, Y; Song, Y. 2021. A bearing fault diagnosis model based on CNN with wide convolution kernels. Journal of Ambient Intelligence and Humanized Computing 13: 4041–4056

Wu, Z.; Wan, Z.; Ge, D. et al. 2022. Car engine sounds recognition based on deformable feature map residual network. Sci Rep 12, 2744 (2022).

Zhou, G; Chen, Y; Chien, C. 2022. On the analysis of data augmentation methods for spectral imaged based heart sound classification using convolutional neural networks. BMC Medical Informatics and Decision Making 22, 226 (2022).

Zhao, Y; Zhang, H; An, L; et al. 2018. Improving the approaches of traffic demand forecasting in the big data era. Cities, Volume 82: 19-26.

**Apêndice ou Anexo** (opcional)

Apêndices são textos e/ou documentos que foram elaborados pelo autor e que são importantes para complementar a argumentação do trabalho. Anexos são textos ou documentos que ilustram, mas que não foram elaborados pelos autores. Apêndices deverão seguir as mesmas normas de formatação do restante do texto, inclusive para figuras e tabelas.

O TCC deverá conter no máximo 30 páginas, incluindo o(s) Apêndice(s) e/ou Anexo(s).