

Instrumentação das Técnicas Árvore de Decisão e KerasTuner para Autoajuste de Redes Neurais Recorrentes GRU na Perspectiva de Computação Sustentável

Davi Lopes Lemos¹, Bianca de Oliveira Durgante¹, Matheus Caneda Correia¹,
Naylor Bastiani Perez², Leonardo Bidese de Pinho¹

¹Universidade Federal do Pampa (UNIPAMPA) - Campus Bagé
96.413-172 – Bagé – RS – Brasil

²Embrapa Pecuária Sul
96.401-970 – Bagé – RS – Brasil

{biancadurgante, davilemos, matheuscorreia}.aluno@unipampa.edu.br

leonardopinho@unipampa.edu.br

naylor.perez@embrapa.br

Resumo. *Este estudo compara o desempenho de duas abordagens para ajuste de hiperparâmetros em modelo de predição com RNN GRU: uma utilizando KerasTuner e outra empregando Árvores de Decisão. A instrumentação foi realizada com a ferramenta Perf para avaliar a eficiência das abordagens perante conjuntos de dados de diferentes dimensões. Os resultados indicam que a combinação de GRU com Árvores de Decisão reduz o tempo de execução em até 30% em comparação à combinação de GRU com KerasTuner, consumindo menos instruções e assim evidenciando-se como uma solução mais eficiente e alinhada aos princípios de computação sustentável.*

1. Introdução

O crescente uso da inteligência artificial (IA) tem impulsionado inovações e avanços em diversos setores. No entanto, essa evolução também intensifica o consumo energético, especialmente devido ao treinamento de modelos complexos em infraestruturas computacionais de alto desempenho. Recentemente, conforme elucidado pela Revista Nature [Science 2023] o setor de Tecnologia da Informação TI ultrapassou o setor de aviação em emissões de Gases de Efeito Estufa, ampliando as discussões na comunidade científica sobre o impacto ambiental do uso dessas tecnologias. Diante desse cenário, cresce a necessidade de soluções que conciliem o progresso tecnológico com a sustentabilidade, mitigando os efeitos ambientais sem comprometer os avanços na área.

Nesse contexto, o presente trabalho propõe a adoção de práticas de computação sustentável, no contexto da redução do tempo de execução, e sua incorporação a um modelo de predição anteriormente proposto [Schulte et al. 2019], denominado Modelo Original (MO), que consiste em uma ferramenta de suporte à decisão e manejo para produtores rurais no âmbito da pecuária extensiva de corte. O estudo baseia-se na integração da abordagem inovadora que utilizou Árvores de Decisão (AD) para mitigar o alto custo associado a ferramentas de autoajuste [Lemos et al. 2024], e da abordagem que avaliou o desempenho da técnica *Gated Recurrent Unit* (GRU) por meio da instrumentação com

a ferramenta oficial do Linux, Perf, visando à otimização do tempo de execução [Durgante et al. 2024]. Dessa forma, busca-se combinar os benefícios das abordagens GRU e AD quando aplicadas ao MO. Além disso, considerando o desafio recorrente do volume de dados necessário para o desempenho satisfatório das RNN, o presente estudo visa uma adaptação para a exploração do conjunto de dados extraído do potreiro 20 infestado (exemplo de área manejada para pecuária, sem nenhum tratamento para plantas indesejáveis), com o objetivo de aprimorar a acurácia do modelo.

2. Materiais e Métodos

Esta pesquisa exploratória aplica modelos que representam aprimoramentos ao sistema de suporte à decisão (MO) que utiliza de Rede Neural Recorrente (RNN) com técnica *Long Short-Term Memory* (LSTM) para prever a quantidade disponível de massa de forragem (MStotal), considerando séries temporais em pastagens naturais do bioma Pampa [Schulte et al. 2019]. O modelo processa entradas meteorológicas e de ambiente coletadas entre janeiro de 2014 e julho de 2018, por meio de uma rede com três camadas de 30, 15 e 1 neurônios, respectivamente. Por padrão, as doze últimas coletas registradas são usadas para validação (equivalente à previsão de um ano), enquanto as demais são empregadas no treinamento.

Os modelos analisados neste trabalho combinam a arquitetura GRU com duas abordagens distintas para autoajuste. A primeira baseia-se em AD e utiliza a biblioteca Python `scikit-learn` para processar um conjunto de dados composto por diferentes combinações de hiperparâmetros [Lemos et al. 2024]. O objetivo é gerar uma estrutura de dados capaz de classificar o ajuste como adequado ou inadequado (considerando um limite de Erro Médio Quadrático [RMSE] de 1700 kg/ha) e reabastecer seu próprio conjunto de dados com os resultados obtidos após o treinamento. A segunda explora o potencial da integração entre GRU e KerasTuner (KT) [O'Malley et al. 2019], utilizando-o como ferramenta para a sintonização do modelo por meio de busca aleatória [Durgante et al. 2024].

Além disso, analisa-se um conjunto de dados base e sua versão expandida, gerada a partir de múltiplas amostras coletadas em subáreas do potreiro 20 infestado. Cada subárea corresponde a uma subdivisão desse potreiro (Figura 1), definida para minimizar as variações nas características do solo e do terreno. Na abordagem inicial (MStotal por potreiro), as amostras coletadas em um mesmo potreiro eram combinadas por meio de uma média ponderada, resultando em um único valor representativo da região. Já na nova abordagem (MStotal por subárea), cada amostra é considerada individualmente, sem a aplicação de proporções, uma vez que não houve impacto positivo na acurácia ao utilizá-las dessa forma. Essa abordagem foi adotada com o objetivo de expandir o conjunto de dados utilizando amostras reais, permitindo a avaliação do impacto dessa expansão no desempenho do modelo. Complementarmente, realizou-se a instrumentação com a ferramenta Perf, considerando eventos de *hardware* como *branch-misses*, *page-faults*, ciclos, instruções e ciclos por instrução (CPI) e para minimizar efeitos de aleatoriedade da rede, cada teste foi repetido dez vezes utilizando o comando `"perf stat -r 10"` [Durgante et al. 2024]. Os experimentos foram realizados em uma máquina local com microprocessador *multicore* AMD Ryzen 7 5700U com frequência de até 4,3 GHz e 8 núcleos com 16 *threads*, placa de vídeo integrada AMD Radeon Graphics 8, 12 GB de memória DDR4 e sistema operacional Ubuntu 22.04.4 LTS, através da versão Python 3.11.5.

3. Resultados e Discussões

Conforme exposto na Tabela 1, AD e KT apresentam valores próximos em relação ao número de falhas de página. No entanto, enquanto AD registra um CPI superior ao de KT, seu número total de instruções é menor, o que justifica a redução no tempo de execução observada na Figura 2. Essa diminuição no volume de instruções processadas impacta diretamente a eficiência computacional do modelo, evidenciando a eficiência da abordagem AD em comparação com KT.

Tabela 1. Resultados dos eventos de *hardware* gerados pelas aplicações

MODELO	<i>Branch-miss</i> (%)	<i>Page-faults</i> (K/s)	<i>Cycles</i> (10 ¹²)	<i>Instructions</i> (10 ¹²)	<i>CPI</i> (Cyc./Ins.)
GRU + AD	4,13	1,96	2,46	0,78	3,15
GRU + KT	3,65	1,996	2,63	1,13	2,31

Adicionalmente, a análise do tempo de execução dos modelos evidencia uma redução significativa no tempo de processamento ao comparar o uso de AD com KT, conforme ilustrado na Figura 2. No entanto, ambos os modelos ainda apresentam um tempo de execução maior em relação ao MO e nesse sentido é importante ressaltar que o ajuste empírico adotado no MO não é contabilizado no tempo total de execução, apesar de demandar intervenção humana e a repetição do processo várias vezes até a obtenção de um ajuste satisfatório dos hiperparâmetros.

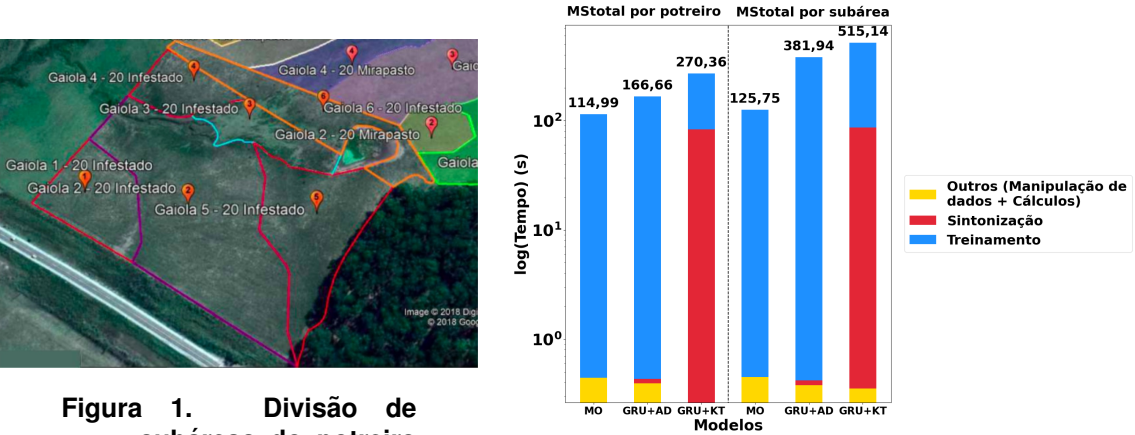


Figura 1. Divisão de subáreas do potreiro 20 infestado

Figura 2. Tempo de execução por modelo e conjunto de dados

A partir da Figura 2, observa-se uma diminuição de aproximadamente 30% no tempo de execução para o conjunto de dados original e de 26% para o expandido, confirmando a hipótese inicial de que a combinação da técnica GRU com AD resultaria em um desempenho superior sem comprometer a acurácia, visto que o tempo de execução para sintonização é substancialmente menor, bem como os treinamentos decorrentes destes ajustes, em ambos conjuntos de dados. Essa abordagem se alinha às práticas de computação sustentável, uma vez que, considerando a necessidade contínua de retreinamento de modelos de IA, a redução do tempo de execução contribui diretamente para a minimização da pegada de carbono associada à solução computacional.

4. Conclusão

Este estudo investigou a combinação da técnica GRU com a abordagem AD para reduzir o tempo de execução de modelos de IA, visando a adoção de práticas de computação sustentável. Os resultados demonstraram que essa estratégia melhora o desempenho do modelo, pois consome menos instruções e, conseqüentemente, reduz o tempo de execução sem comprometer a acurácia das previsões, mesmo quando o conjunto de dados é ampliado com novas amostras. Dessa forma, sua utilização possibilita uma economia significativa de consumo energético, alinhando-se aos objetivos de sustentabilidade e eficiência energética na execução de modelos preditivos. Destaca-se ainda o impacto dessa redução de tempo de execução, considerando que modelos de IA precisam ser constantemente retreinados para manter sua eficiência.

Em suma esta pesquisa visa contribuir com o estado da arte incorporando *Smart Farming*, tecnologias modernas na agricultura digital, pecuária de precisão e aspectos Ambientais, Sociais e de Governança (ESG). No futuro almeja-se explorar com maior profundidade a questão da expansão de dados através das subáreas no que tange a ponderação das amostras. Ademais, salienta-se a indissociabilidade entre Ensino, Pesquisa e Extensão na obtenção e análise dos resultados, uma vez que conceitos de disciplinas da EC foram diretamente aplicados, discutidos e revisitados em um problema conectado com o contexto regional do Bioma Pampa [Pinho and Domingues Júnior 2024].

Agradecimentos

Reconhece-se o apoio da CAPES - Código de Financiamento 001 (PROAP) e dos Programas institucionais de Bolsas de iniciação científica PIBIC/CNPq, PROBIC/FAPERGS e PRO-IC/UNIPAMPA.

Referências

- Durgante, B. O., Lemos, D. L., Perez, N. B., and Pinho, L. B. (2024). Análise instrumental de modelos rnn para agricultura digital visando computação verde. *Anais do Salão Internacional de Ensino, Pesquisa e Extensão*, 2(16). Acesso em: 3 fev. 2025.
- Lemos, D. L., Durgante, B. O., Perez, N. B., and Pinho, L. B. (2024). Árvores de decisão como solução para ajuste de modelo de previsão para agricultura digital. In *Encontro de Iniciação Científica e Tecnológica do IFSul (ENCIF)*, Bagé, RS, Brasil.
- O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al. (2019). Kerastuner. <https://github.com/keras-team/keras-tuner>. [online: acesso em 24-janeiro-2024].
- Pinho, L. B. and Domingues Júnior, J. S. (2024). Enhancing computer architecture learning with a spiral curriculum. *International Journal of Computer Architecture Education*, 13(1):61–70.
- Schulte, L. G., Perez, N. B., Pinho, L. B., and Trentin, G. (2019). Sistema de apoio à decisão para pecuária de precisão: Módulo para ajuste de taxa de lotação. In *Anais do XV Simpósio Brasileiro de Sistemas de Informação (SBSI)*, pages 95–102, Porto Alegre, RS, Brasil. SBC.
- Science, N. C. (2023). The carbon footprint of computational research. *Nature Computational Science*, 3:659.