

Fine-tuning e Análise do Aprendizado de Modelos de Linguagem Navegando por Labirintos em ASCII

Anderson N. Silva¹

¹Departamento Acadêmico de Informática (DAINF) –
Universidade Tecnológica Federal do Paraná (UTFPR)
Av. Sete de Setembro, 3165 – 80230-901 – Curitiba – PR – Brasil

`andersonnogueira@alunos.utfpr.edu.br`

Abstract. *10 lines and must be in the first page of the paper.*

Resumo. *Este meta-artigo descreve o estilo a ser usado na confecção de artigos e*

1. Introdução

2. Fundamentos e Trabalhos Relacionados

Esta seção discute os trabalhos relacionados e os fundamentos teóricos principais relacionados com o tema proposto.

2.1. LLMs - Large Language Models

Os modelos de Linguagem de Grande Escala (*Large Language Models - LLMs*) são arquiteturas computacionais baseadas em redes neurais profundas, predominantemente do tipo *transformer*, desenvolvidos para processar e gerar linguagem natural. Esses modelos são treinados com grandes volumes de dados textuais, com os quais conseguem (computar/aprender) padrões estatísticos da linguagem humana, tornando-os capazes de realizar uma ampla gama de tarefas, grande parte sem treinamento supervisionado específico (e.g. sumarização, resposta a perguntas e tradução). A característica comumente associada ao tamanho desses modelos é uma medida em bilhões de parâmetros que está diretamente associado à sua capacidade de adaptação a contextos diversos e sua generalização. A arquitetura *transformer*, introduzida por [Vaswani et al. 2023], faz parte da estrutura central da maioria dos LLMs atuais, e destaca-se pelo uso de mecanismos de atenção que possibilitam a modelagem de dependências contextuais de longo alcance.

2.2. Fine-tuning

O *fine-tuning* é uma técnica utilizada para adaptar *LLMs* a tarefas específicas, por meio de re-treinamento de parte ou totalidade de seus parâmetros com conjuntos de dados mais específicos. Em seguida da etapa inicial de pré-treinamento, na qual um *LLM* aprende as representações gerais linguagem a partir de grandes corpora, o *fine-tuning* permite especializar o comportamento do modelo para contextos mais delimitados como, por exemplo, geração de código, classificação de sentimentos ou, como no presente trabalho, navegação por labirinto em caracteres ASCII. Existem diferentes estratégias de *fine-tuning*:

- Supervised Fine-Tuning (SFT): técnica de ajuste fino supervisionado, onde o modelo é treinado com um conjunto de dados rotulado, permitindo que aprenda a mapear entradas para saídas específicas. Essa abordagem é comum para tarefas como classificação, tradução e resposta a perguntas.

- Direct Preference Optimisation (DPO): técnica de otimização direta de preferências, onde o modelo é ajustado para maximizar a probabilidade de gerar respostas preferidas por humanos, com base em feedback direto. Essa abordagem é útil para melhorar a qualidade das respostas geradas pelo modelo.
- Reinforcement Learning from Human Feedback (RLHF): técnica de aprendizado por reforço a partir de feedback humano, onde o modelo é treinado para maximizar uma recompensa baseada em preferências humanas. Essa abordagem é eficaz para melhorar a qualidade das respostas geradas pelo modelo, alinhando-as com as expectativas dos usuários.
- Odds Ratio Preference Optimization (ORPO): técnica de otimização de preferências baseada em razão de chances, onde o modelo é ajustado para maximizar a razão entre as probabilidades de gerar respostas preferidas e não preferidas. Essa abordagem é útil para melhorar a qualidade das respostas geradas pelo modelo, alinhando-as com as expectativas dos usuários.
- Group Relative Policy Optimization (GRPO): técnica de otimização de políticas relativas em grupo, onde o modelo é ajustado para maximizar a probabilidade de gerar respostas preferidas por um grupo de usuários, levando em consideração as preferências coletivas. Essa abordagem é eficaz para melhorar a qualidade das respostas geradas pelo modelo, alinhando-as com as expectativas de um público mais amplo.

2.3. Explicabilidade em LLMs

Apesar de possuírem um alto nível de compreensão e resolução daquilo que lhes são atribuídos, os LLMs são questionados com frequência sobre seu caráter de “caixa-preta”, que é a dificuldade no entendimento sobre como tomam decisões ou geram suas respostas.

Nesse contexto, é cada vez maior o interesse em técnicas de explicabilidade (*explainability*) aplicadas a redes neurais profundas para tornar seus mecanismos internos interpretáveis. A explicabilidade em LLMs tem o objetivo de revelar como os dados são processados ao longo das camadas do modelo, quais neurônios foram ativados em resposta a determinadas entradas, como essas ativações influenciam a saída final e como as representações internas evoluem durante o treinamento ou *fine-tuning*.

Essas técnicas são essenciais para o diagnóstico de falhas, enviesamento e limitações de raciocínio do modelo. Diversas ferramentas vêm sendo propostas para auxiliar na explicabilidade de LLMs com visualizações, utilização de grafos de conhecimento, análise de atenção e outras abordagens que ajudam a decifrar o comportamento dos modelos.

2.3.1. LLM-MRI

A biblioteca LLM-MRI (*Large Language Model - Magnetic Resonance Imaging*) é uma das ferramentas desenvolvidas com o intuito de auxiliar na explicabilidade de LLMs. Ela atua com o objetivo de simplificar o estudo de padrões de ativações em qualquer LLM baseado em *transformer*.

Desenvolvida por [Costa et al. 2024], a LLM-MRI permite a coleta, organização e projeção em dimensões reduzidas dos vetores de ativação produzidos pelos modelos

ao processarem diferentes entradas. Por meio de técnicas de redução de dimensionalidade, é possível visualizar como as ativações neuronais se distribuem em um espaço de menor dimensão, facilitando a identificação de padrões e também como as ativações se diferenciam após a aplicação de diferentes técnicas de fine-tuning.

Essa abordagem é relevante, considerando o contexto experimental deste trabalho, avaliar como LLMs se adaptam a tarefas específicas (neste caso, navegação em labirintos ASCII) e como as estruturas internas dos modelos emergem a partir dessa aprendizagem.

2.4. Trabalhos Relacionados

Esta seção apresenta a revisão de literatura sobre o tema deste trabalho, com foco em LLMs, fine-tuning e explicabilidade.

2.4.1. Raciocínio Espacial em LLMs

Entre as abordagens relevantes envolvendo o raciocínio espacial de LLMs, destaca-se o trabalho de [Dao and Vu 2025], que introduz o framework AlphaMaze como uma estratégia, em duas etapas, para capacitar LLMs a resolverem labirintos representados em texto. Inicialmente, os modelos passam por um Supervised Fine-Tuning (SFT) com dados tokenizados para aprender comandos de movimentação passo a passo; em seguida, aplica-se o algoritmo Group Relative Policy Optimization (GRPO) para promover um raciocínio mais robusto e autocorretivo. O estudo apresenta também o MazeBench, usado para avaliar o desempenho dos modelos com labirintos de diferentes níveis de dificuldade, e demonstra um incremento significativo de desempenho após a aplicação do algoritmo GRPO.

De modo complementar, estudos como o de [Jiang et al. 2024] questionam a capacidade de raciocínio espacial em modelos multimodais, mostrando que modelos que processam apenas entradas textuais podem superar modelos multimodais em tarefas de raciocínio espacial. Para tarefas espaciais mais rigorosas, como o uso de LLMs para entendimento de espaços tridimensionais estruturados, [Zhang et al. 2025] demonstrou que, apesar dos avanços significativos em outras áreas, LLMs ainda enfrentam desafios ao processar tais relações. Paralelamente, [Cheng et al. 2024] desenvolve um framework que visa aprimorar a percepção e raciocínio tridimensional em modelos de linguagem visual, integrando módulos de representação de regiões e plugins para informações de profundidade, permitindo uma compreensão mais profunda das relações espaciais em cenas complexas.

Diferentemente de outras abordagens que requerem fine-tuning adicional, [Hu et al. 2023] introduz o método Chain-of-Symbol (COS) Prompting como uma técnica inovadora para aprimorar a capacidade de raciocínio espacial em LLMs. O COS Prompting converte descrições em linguagem natural para representações simbólicas durante as etapas intermediárias de raciocínio, permitindo que o modelo utilize símbolos de forma mais eficiente. Os resultados experimentais mostraram ganhos significativos em tarefas de planejamento espacial, incluindo reduções expressivas no uso de tokens e melhorias substanciais na acurácia em benchmarks específicos.

O método intitulado Visualization of Thought (VoT) também se destaca como uma abordagem alternativa para aprimorar o raciocínio espacial em LLMs. Inspirando-

se na cognição humana, [Wu et al. 2024] propõe o VoT como uma técnica que estimula a criação de imagens mentais por parte dos modelos utilizando prompts específicos que guiam a geração de representações visuais em ASCII art durante a execução de tarefas, como resolução de labirintos e planejamento de caminhos. Experimentos realizados com o GPT-4 demonstraram que o VoT melhora a capacidade do modelo em acompanhar estados e planejar ações de forma mais precisa, superando abordagens tradicionais e modelos multimodais em determinados casos. Contudo, os resultados também indicam que essa abordagem depende fortemente das habilidades emergentes dos modelos e da qualidade dos prompts, mostrando limitações em cenários mais complexos ou com modelos menos robustos.

2.4.2. Fine-tuning de LLMs para Tarefas Específicas

O fine-tuning de LLMs para tarefas específicas é um tema amplamente investigado, com diferentes estratégias sendo propostas para aprimorar a adaptação dos modelos. Por exemplo, [Wang et al. 2024] demonstra que o uso de SFT pode melhorar tanto as representações internas quanto a capacidade de generalização, mesmo em modelos visuais, evidenciando a importância do ajuste direcionado para tarefas de raciocínio espacial. Além disso, [Hsieh et al. 2023] propõem a utilização de rationales gerados por LLMs como supervisão adicional em um cenário multitarefa, combinando rótulos e explicações para enriquecer o sinal de treinamento. Essa abordagem, baseada em Chain-of-Thought prompting [Wei et al. 2023], permite que modelos menores alcancem desempenho superior a LLMs maiores, mesmo com menos dados, ao tornar o processo de ajuste mais eficiente e interpretável. No contexto deste trabalho, a integração de racionalizações se mostra relevante para aprimorar a eficiência e a interpretabilidade dos modelos em tarefas como a resolução de labirintos ASCII.

2.4.3. Explicabilidade em LLMs

A explicabilidade em LLMs é um campo em expansão, com abordagens que buscam tornar os modelos mais transparentes e compreensíveis. A biblioteca LLM-MRI [Costa et al. 2024] destaca-se por permitir a análise e visualização das ativações neuronais, facilitando a interpretação de como os modelos se adaptam a tarefas específicas, como a navegação em labirintos ASCII. Complementarmente, [Wu et al. 2025] propõem o conceito de Usable XAI, que enfatiza a aplicação prática das explicações para diagnóstico e aprimoramento dos modelos. Entre as estratégias discutidas estão métodos de atribuição, análise das ativações internas e explicações baseadas em exemplos, todas relevantes para entender a evolução das representações internas durante o fine-tuning. A integração dessas abordagens contribui para identificar limitações, corrigir vieses e aprimorar o desempenho dos modelos em tarefas que exigem alto grau de interpretabilidade.

References

- Cheng, A.-C., Yin, H., Fu, Y., Guo, Q., Yang, R., Kautz, J., Wang, X., and Liu, S. (2024). Spatialrgpt: Grounded spatial reasoning in vision language models. *arXiv preprint arXiv:2406.01584*.

- Costa, L., Figênio, M., Santanchè, A., and Gomes-Jr, L. (2024). LLM-MRI Python module: a brain scanner for LLMs. In *Anais Estendidos do XXXIX Simpósio Brasileiro de Bancos de Dados*, pages 125–130, Porto Alegre, RS, Brasil. SBC.
- Dao, A. and Vu, D. B. (2025). Alphamaze: Enhancing large language models’ spatial intelligence via grpo. *arXiv preprint arXiv:2502.14669*.
- Hsieh, C.-Y., Li, C.-L., Yeh, C.-K., Nakhost, H., Fujii, Y., Ratner, A., Krishna, R., Lee, C.-Y., and Pfister, T. (2023). Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes.
- Hu, H., Lu, H., Zhang, H., Song, Y.-Z., Lam, W., and Zhang, Y. (2023). Chain-of-symbol prompting elicits planning in large language models. *arXiv preprint arXiv:2305.10276*.
- Jiang, X., Ge, Y., Ge, Y., Shi, D., Yuan, C., and Shan, Y. (2024). Supervised fine-tuning in turn improves visual foundation models. *arXiv preprint arXiv:2401.10222*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2023). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- Wang, J., Ming, Y., Shi, Z., Vineet, V., Wang, X., Li, S., and Joshi, N. (2024). Is a picture worth a thousand words? delving into spatial reasoning for vision language models. *Advances in Neural Information Processing Systems*, 37:75392–75421.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., and Zhou, D. (2023). Chain-of-thought prompting elicits reasoning in large language models.
- Wu, W., Mao, S., Zhang, Y., Xia, Y., Dong, L., Cui, L., and Wei, F. (2024). Mind’s eye of llms: Visualization-of-thought elicits spatial reasoning in large language models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.
- Wu, X., Zhao, H., Zhu, Y., Shi, Y., Yang, F., Hu, L., Liu, T., Zhai, X., Yao, W., Li, J., Du, M., and Liu, N. (2025). Usable xai: 10 strategies towards exploiting explainability in the llm era.
- Zhang, W., Peng, R., Gao, C., Fang, J., Zeng, X., Li, K., Wang, Z., Cui, J., Wang, X., Chen, X., et al. (2025). The point, the vision and the text: Does point cloud boost spatial reasoning of large language models? *arXiv preprint arXiv:2504.04540*.