Language Models Explore the Linguistics of Chess

Lars Schmid and **Jerome Maag** and **Mark Cieliebak** and **Pius von Däniken**
schmila7@students.zhaw.ch

Abstract

In this research, we explore how to train Large Language Models (LLMs) to generate valid chess moves, extending their use from traditional Natural Language Processing to structured, rule-based language systems. We trained the openly available GPT-2 architecture on chess notation datasets containing up to one million games. We then evaluated model performance using the following metrics: "Average Number of Correct Plies", which assessed the model's ability to generate moves for both players by recording the average number of correct moves across 100 simulated games; "Hard Position Accuracy," which evaluated the model's capability to handle 67 challenging positions, assessing the percentage of scenarios successfully solved; and "Legal Piece Moves Accuracy," which specifically evaluated the model's ability to generate valid moves for a given piece, based on its position on the board. This last metric also aimed to assess the model's capability in modelling the state of the game board. The evaluations revealed a significant correlation between dataset size and model effectiveness, particularly emphasizing the importance of integrating the Beginning of Sequence (BOS) token to enhance syntactical correctness and reduce errors. Despite challenges with some complex chess positions, the best model achieved an average accuracy rate of up to 50 correct moves per game, demonstrating the adaptability of LLMs to rule-based systems like chess and opening new possibilities for AI in structured domains.