Information Retrieval Project Proposal

Tobias Wilfert [s0181439], John Castillo [s0175247]

Scope

Since we are both avid chess players since our childhood with thousands of games played, we decided that we wanted to work on a chess-related topic. Looking at the example projects we got intrigued by the Related Document Retrieval. We realized that the state-of-the-art chess analysis tool let you look for exact position matches in a database but not similar positions. Looking at similar positions, however, becomes interesting during mid-game scenarios where not all pieces are relevant to the current position. Many of these mid-game scenarios are completely unique yet share the same essence. Since they share the same essence, you can learn by seeing how these positions unfold. This relates closely to how one learns from his own games where the positions might not always be the exact same, but the ideas are. However, finding these games with similar positions is currently almost impossible. As such we thus propose a chess explorer that lets players specify a position to query similar positions in a chess database.

Resources

We are planning on using a subset of the Lichess open database (https://database.lichess.org) which consists of 4 billion standard games played on Lichess. There is also a subset available of the games played by elite chess players (2400+ ELO) which might be interesting. The games are represented as the moves played during the game plus metadata about the players and the game itself.

For the programming language, we are planning on using Python in combination with PyLucene. Since most chess positions are unique the lookup will rely on ranked retrieval. For storing the positions, we are thinking about using the popular Forsyth–Edwards Notation although we might need to tweak it.

Evaluation

The evaluation is slightly tricky due to the lack of ground truth. To mitigate this, we plan to query positions similar to game-states we know are in the Database to check that they appear as a search result. Furthermore, we plan to run multiple queries and reason about the output our program produces.

Future work

If we manage to reliably look up positions with the same essence this could then be used to solve chess puzzles. These consist of a specific position in which the player needs to find the best next move. A player could then look at similar positions and how the games panned out to decide what the best move is. Lichess also has a Database for Puzzles consisting of 3 million puzzles together with their solutions.

Literature

We are planning on reading the following documents as we deem aspect of them interesting or related to our project:

- *Debasis Ganguly, Johannes Leveling, and Gareth J.F. Jones. 2014. Retrieval of similar chess positions. In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (SIGIR '14). Association for Computing Machinery, New York, NY, USA, 687–696. https://doi.org/10.1145/2600428.2609605*

- *Bizjak, M., Guid, M. (2022). Automatic Recognition of Similar Chess Motifs. In: Browne, C., Kishimoto, A., Schaeffer, J. (eds) Advances in Computer Games. ACG 2021. Lecture Notes in Computer Science, vol 13262. Springer, Cham. https://doi.org/10.1007/978-3-031-11488-5_12*

- *William G. Chase, Herbert A. Simon, Perception in chess, Cognitive Psychology, Volume 4, Issue 1, 1973, Pages 55-81, ISSN 0010-0285, https://doi.org/10.1016/0010-0285(73)90004-2. (https://www.sciencedirect.com/science/article/pii/0010028573900042)*

- *Atsushi Ushiku, Shinsuke Mori, Hirotaka Kameko, and Yoshimasa Tsuruoka. 2017. Game State Retrieval with Keyword Queries. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17). Association for Computing Machinery, New York, NY, USA, 877–880. https://doi.org/10.1145/3077136.3080668*

- *(Potentially interesting) https://content.iospress.com/articles/icga-journal/icg21204*