Move-by-Move Commentary for Chess Games

Gavrilyuk Alexander

May 2024

Abstract

This paper gives a brief overview of the problem of generating natural language descriptions of chess games and describes my attempt to contribute in the process of solving it.

Project code is located here: `https://github.com/CrubBucket/NLP_chess_commentary`.

1 Introduction

The problem of automatic chess games descriptions generation is not quite popular among non-chess community, but is crucial for chess enthusiasts. Having a tool to generate high-quality comments for each move in a chess game would make an education process easier and more pleasant for players. Nowadays we have very strong chess engines which can easily beat any grand-master but studying lines these engines provide can be quite complicated because of the lack of explanations. So, NLP-approach can fix it and add more humanity to raw machine analysis.

1.1 Team

Gavrilyuk Alexander

2 Related Work

There are not so many papers and works related to this topic due to its unpopularity. One of the most detailed and resent ones is a paper "Learning to Generate Move-by-Move Commentary for Chess Games from Large-Scale Social Forum Data"[1]

First of all, authors introduce a new large-scale chess commentary dataset consists of 298K aligned game move/commentary pairs tables 1, 2. The dataset D consists of data points of the form $(Si, Mi, Gi), i \in \{1, 2, .., |D|\}$ where S_i is the commentary text for move M_i and G_i is the corresponding chess game. S_{i_1} is a sequence of m tokens $S_{i_1}, S_{i_2}, ..., S_{i_m}$, so the solution of the move commentary

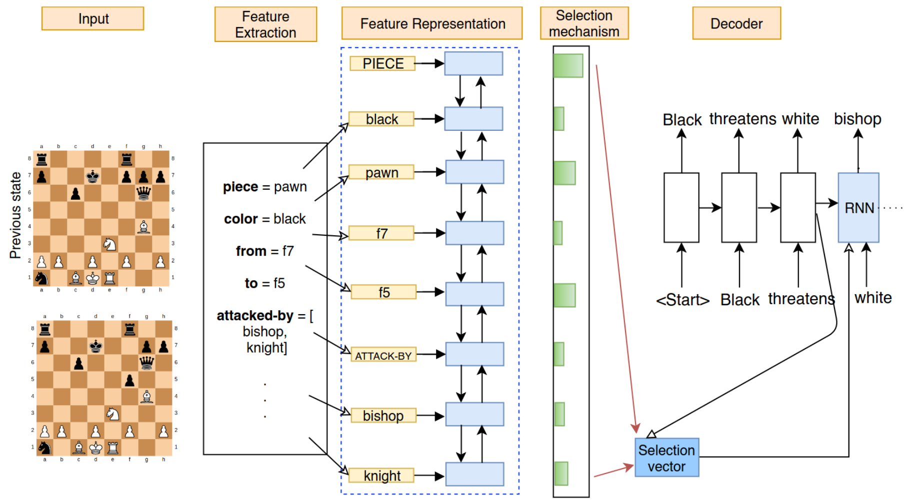

Figure 1: A model overview.

boils down to modeling the probability $P(S_i|M_i, C_i, R_i)$, where C_i and Ri are current and previous board stages.

Authors consider the following features to provide their model with necessary information to generate commentary texts:

- **Move** features $f_{move}(M_i, C_i, R_i)$ - encode the general current move information
- **Threat** features $f_{threat}(M_i, C_i, R_i)$ - encode information about pieces of opposite player attacking the moved piece before and after the move
- **Score** features $f_{score}(M_i, C_i, R_i)$ - capture the quality of move and general progress of the game

Representing of all these features is considered through embeddings. For categorical features authors directly look up the embedding using corresponding token, while real valued features are encoded by first binning them and then using corresponding number for embedding lookup. Let E represent the embedding matrix. Then $E[f^j_{move}]$ represents embeddings of j_{th} move feature, or in general $E[f_{move}]$ represents the concatenated embeddings of all move features. Similarly, $E(f_{move}, f_{threat}, f_{score})$ represents concatenated embeddings of all the features. Finally, authors' model contain LSTM as a decoder. At every output step, the LSTM decoder predicts a distribution over vocabulary words to construct a commentary output. Figure 1 demonstrates a schematic model overview.

Authors mainly observe BLEU and BLEU-2 scores to measure the performance of the models and conduct a human evaluation study (tables 3 and 4)

Dataset stats	
Total Games	11,578
Total Moves	298,008
Average no. of recorded steps in a game	25.73
Frequent Word Types	39,424
Rare Word Types	167,321
Average Comment Length	20.55
Long Comments	230745 (77%)

Table 1: Dataset and Vocabulary Statistics

Category	**Example**	**% in data**	**Val acc.**
Direct Move Description	An attack on the queen	31.4%	71%
Move Quality	A rook blunder.	8.0%	90%
Comparative	At this stage I figured I better move my knight.	3.7%	77.7%
Planning / Rationale	Trying to force a way to eliminate d5 and prevent Bb5.	31.2%	65%
Contextual Game Info	Somehow, the game I should have lost turned around in my favor.	12.6%	87%
General Comment	Protect Calvin , Hobbs	29.9%	78%

Table 2: Commentary classification

Features	**BLEU**	**BLEU-2**
M	1.69	20.66
M+T	1.94	24.11
M+T+S	2.02	24.70

Table 3: Results of main model learned on the whole training set using different combinations of features (M - move features, T - threat features, S - score features)

Question	GT	GAC (M)	GAC (M+T)	GAC (M+T+S)
Is commentary correct for the given move? (%Yes)	70.4	42.3	64.8	67.6
Can the move be inferred from the commentary? (%Yes)	45.1	25.3	42.3	36.7
Fluency (scale of (least)1 - 5(most)) Mean (Std. dev.)	4.03 (1.31)	4.15 (1.20)	4.44 (1.02)	4.54 (0.89)

Table 4: Human study results (GT - ground truth, GAC - main model)

3 Model Description

Unfortunately, I did not manage to finish this project during the course, but here are some thoughts on how I would modify the architecture given above:

- Using GRU instead of LSTM (at least, if they perform similarly, we could reduce the computational expenses).
- Extending the feature types extracted from different board stages.
- Considering more chess board stages (not only current and previous ones).
- Changing strateges for calculating embeddings.
- Experimenting with ensembling separately trained models.

4 Dataset

As it was mentioned above, there is a lack of works related to this topic due to its unpopularity, so the main dataset used in this project is the one from previously described paper. It can be collected with the open-source code[2] provided by authors (as it is totally available for the research purposes).

It is decided to enrich the given dataset with some additional data, which was mined through parsing open Lichess database. 5.7k moves of 56 games of Candidates Tournament 2024 followed by professional commentary were parsed. Presumably, top level chess content could improve the dataset quality and give some boost to model results. For example, a new category "Professional commentary" can be added to the list of comments classes.

On the Tab. 5 you can see the statistics for the mentioned dataset.

Additional dataset contain move notation, FEN notation (to extract all the features connected with pieces placement), current engine evaluation of the position after the certain move and a comment itself.

Dataset stats	
Total Games	56
Total Moves	5665
New words	3207
Average Comment Length	23.06

Table 5: Additional Dataset and Vocabulary Statistics

5 Conclusion

This work is my starting point of a big NLP-project dedicated to automatic chess games commentary generation. Unfortunately, only related works have been studied and a strategy to collect new data and extend existing dataset have been formulated, but the project will undergo further development. If you are interested in this topic, you can spectate this work's growth or even contribute to it via the repository: `https://github.com/yournickname/your-project-name`

References

[1] Harsh Jhamtani Varun Gangal Eduard Hovy Graham Neubig Taylor Berg-Kirkpatrick. "Learning to Generate Move-by-Move Commentary for Chess Games from Large-Scale Social Forum Data". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (2018), pp. 1661–1671. DOI: `https://doi.org/10.18653/v1/P18-1154`.

[2] Harsh Jhamtani et al. "Learning to Generate Move-by-Move Commentary for Chess Games from Large-Scale Social Forum Data". In: *The 56th Annual Meeting of the Association for Computational Linguistics (ACL)*. Melbourne, Australia, July 2018.