Chess Q&A : Question Answering on Chess Games

Volkan Cirik, Louis-Philippe Morency, Eduard Hovy
Department of Computer Science
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213
{vcirik,morency,hovy}@cs.cmu.edu

Abstract

We introduce a new dataset[1] for the evaluation of models addressing reasoning tasks. For a position in a chess game, we provide question and answer pairs, an image of the board, and the sequence of moves up to that position. We hope this synthetic task will improve our understanding in memory based Deep Learning with posed challenges.

1 Introduction

Deep Neural architectures have successfully been trained for numerous tasks such as speech, computer vision and natural language processing [13, 10]. Recent research focuses on extending the capabilities of these architectures to solve deeper reasoning problems.

To address reasoning tasks, an architecture requires attention and memory. Graves et. al. introduce memory and controller modules to handle memory operations for synthetic tasks such as copying a sequence [7]. Weston et. al. propose a memory component for natural language question answering [17]. Attention mechanisms are shown to be successful in image classification [14], machine translation [2], speech recogniton [4] and image captioning [18].

Synthetic datasets play a crucial role in understanding and developing complex machine learning algorithms [16]. To this end, we propose a Chess Q&A, a new dataset for answering questions for a given chess match configurations. Unlike real world tasks, in chess, a limited amount of knowledge is required to answer factual questions. We believe chess Q&A poses new challenges for novel Deep Architectures and help improve their capabilities.

2 Chess Question Answering

Chess is a two-player, board game played on 64 squares arranged on an 8x8 grid. Each player starts with 16 pieces with 6 types of: Pawn, Knight, Bishop, Rook, Queen and King. Each piece has their own set of moves. We suggest further reading[2] for the rules of the game.

Our dataset consists of set of questions about the basic rules of the game. These questions do not require background knowledge further than this. All the questions are noiseless, and a human or a short computer program could solve them with perfect accuracy.

2.1 Question Types

We provide different question types to test the various properties of a chess board and rules.

[1] http://www.cs.cmu.edu/˜vcirik/chess_qa.html
[2] https://en.wikipedia.org/wiki/Chess

2.1.1 Position of a Piece

The first question ask what kind of piece is on given position see Figure 1a for an example. Here, the algorithm should identify the piece either using the inner spatial representation from the image of the board, or its simulated representation of the board configuration from the sequence of moves.

2.1.2 Counting All Pieces

The second question asks for the number of pieces on the board. This will require a counting operation over all the pieces (Figure 1b).

2.1.3 Counting Pieces For A Side

The third type of questions ask for the number of pieces for a given side. This will require grounding the meaning of sides and using a count operation over the pieces (Figure 1c).

2.1.4 Existence of A Piece

This question asks for whether a piece is on the board or not. This will require an existence operation similar to counting (Figure 1d).

2.1.5 Existence of A Piece For A Side

Similar to the previous questions, these ask whether a specified piece is on the board for given side. This will also require grounding the meaning of side (Figure 1e).

2.1.6 Legal Move

These questions will test the rules regarding the movement for each piece. Note that special moves like En Passant and Castling require knowledge of previous moves (Figure 1f).

2.1.7 Attacking a Square

These questions will test the concept of attacking a square.(Figure 1g).

2.1.8 Being Under Attack

These questions ask whether a given square is under attack by a certain piece. This is the inverse of the previous questions. (Figure 1h).

2.1.9 Check

These questions ask whether a given side is in check. This is a special case where the King is the piece being attacked (Figure 1i).

2.1.10 Material Count

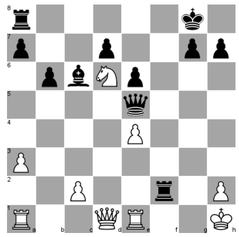
Each type of piece is worth different points during the game. We use a material scale described in [8] where Pawns are worth 1 point, Knights and Bishops are worth 3, Rooks are worth 5, and the Queen is worth 10. These questions ask about the relative material points for a side. (Figure 1j).

2.1.11 Material Advantage

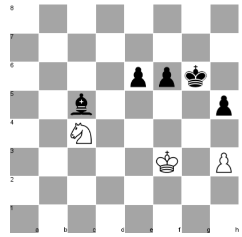
This question asks which side is leading the game. To answer this, the model has to learn the sign of the material count (Figure 1k).

2.1.12 Castling Rights

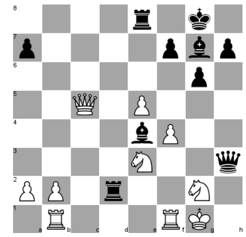
These questions ask whether a given side has castling rights. This requires the model to understand the sequence of moves played, because a King cannot castle if it has previously moved. (Figure 1l).



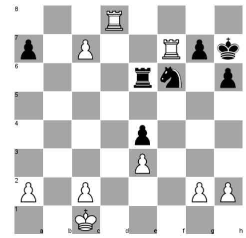
(a) **Q:**What piece is on d6?
A:white knight



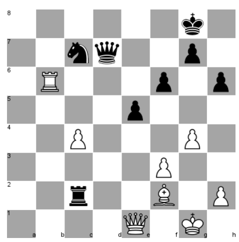
(b) **Q:**How many pieces are on the board?
A:8



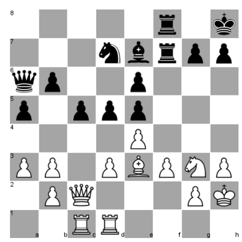
(c) **Q:**How many pieces does black have?
A:10



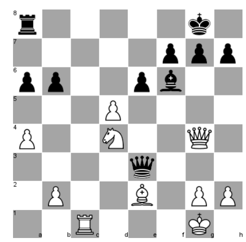
(d) **Q:**Is there any queen on the board?
A:No



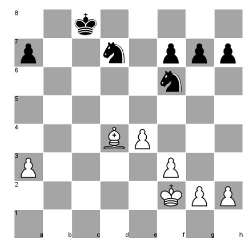
(e) **Q:**Does white have a knight?
A:No



(f) **Q:**Is b2b3 a legal move?
A:No

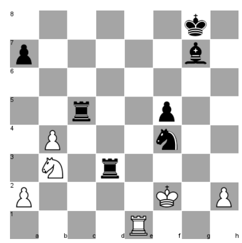


(g) **Q:**Which piece is attacking black pawn at a6?
A:white bishop

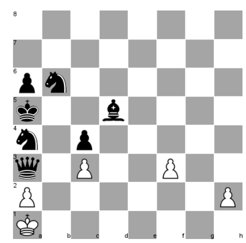


(h) **Q:**Is a7 under attack by white bishop?
A:Yes

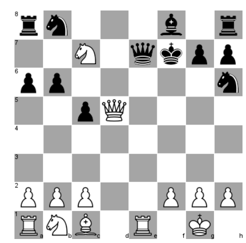
(i) **Q:**Is black in check?
A:Yes



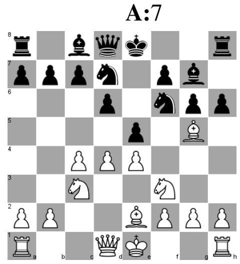
(j) **Q:**What is the material advantage of black?
A:7



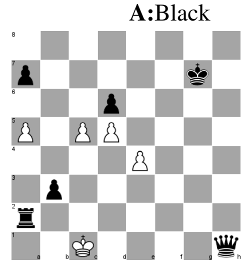
(k) **Q:**Who has the material advantage?
A:Black



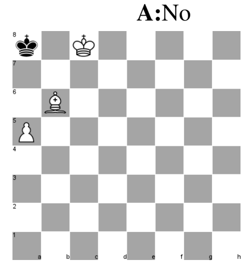
(l) **Q:**Does black has castling rights?
A:No



(m) **Q:**Can white castle?
A:Yes



(n) **Q:**Is this a checkmate?
A:Yes



(o) **Q:**Is this a stalemate?
A:Yes

Figure 1: Examples of each types of questions.

2.1.13 Possible Castling

In addition to the previous question, we ask whether a side can castle for a given board configuration. The side has to have castling rights, the squares between a Rook and King have to be empty, all of the squares in between the Rook and King cannot be under attack, and the King cannot be in check. (Figure 1m).

2.1.14 Checkmate

We ask whether a given configuration is a checkmate. This requires reasoning about whether certain squares are under attack and the possible moves of the King. (Figure 1n).

2.1.15 Stalemate

In stalemate, a given side is not in check but there are no valid moves that can be made. This question is similar to checkmate, however the turn has to be on the side that has no valid moves (Figure 1o).

2.2 Question Preparation

We downloaded games from a Chess game archive[3], and we used blitz games played in January 2014. We generated questions and answers automatically. Each question type has two paraphrases to vary natural language input. Answers could be yes, no, numbers, and pieces names. For each type of question we generated 1000 samples. We made sure that the answer types are balanced (e.g the number of yes/no answers are the same). We provide questions, answers, an image of the board and the sequence of moves in Portable Game Notation[4].

3 Conclusion

We propose a new dataset for question answering. The task is analog to bAbI [16] in the sense that each move (statements) introduces a fact about the game (story). Our dataset is also a synthetic version of visual question answering datasets [11, 5, 1, 18]. Notably, [1] introduces visual question-answering on synthetic Abstract Scene Datasets [19, 20].

We believe the questions we raised from this dataset will be useful in investigating the limitations and capabilities of new architectures. Real world tasks require large amounts of background knowledge and annotation effort. Here, in the limited world of Chess Q&A, we can investigate whether an algorithm can learn (our/a) knowledge base, i.e. the rules of the game. We can also explore whether it is possible to use an existing knowledge base to answer questions.

The proposed dataset can be used in the context of a grounding problem [6, 12, 9, 15]. The word tokens such as piece names, side, positions, sequence of moves, and rules can be grounded on visual representations of the board.

Another research question is whether we can do curriculum learning [3] in this setup. For instance checking (Section 2.1.14) requires the knowledge of being under attack (Section 2.1.8) and legal moves (Section 2.1.6). We hope further research will answer the questions we raised here and in the dataset.

References

[1] Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh. Vqa: Visual question answering. *arXiv preprint arXiv:1505.00468*, 2015.

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[3] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48. ACM, 2009.

[4] Jan Chorowski, Dzmitry Bahdanau, Dmitriy Serdyuk, Kyunghyun Cho, and Yoshua Bengio. Attention-based models for speech recognition. *arXiv preprint arXiv:1506.07503*, 2015.

[5] Haoyuan Gao, Junhua Mao, Jie Zhou, Zhiheng Huang, Lei Wang, and Wei Xu. Are you talking to a machine? dataset and methods for multilingual image question answering. *arXiv preprint arXiv:1505.05612*, 2015.

[3] http://ficsgames.org/download.html

[4] http://www.saremba.de/chessgml/standards/pgn/pgn-complete.htm

[6] Peter Gorniak and Deb Roy. Grounded semantic composition for visual scenes. *Journal of Artificial Intelligence Research*, pages 429–470, 2004.

[7] Alex Graves, Greg Wayne, and Ivo Danihelka. Neural turing machines. *arXiv preprint arXiv:1410.5401*, 2014.

[8] David Hooper and Kenneth Whyld. The oxford companion to chess. *Oxford University Press*.

[9] Jayant Krishnamurthy and Thomas Kollar. Jointly learning to parse and perceive: Connecting natural language to the physical world. *Transactions of the Association for Computational Linguistics*, 1:193–206, 2013.

[10] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[11] Mateusz Malinowski, Marcus Rohrbach, and Mario Fritz. Ask your neurons: A neural-based approach to answering questions about images. *arXiv preprint arXiv:1505.01121*, 2015.

[12] Cynthia Matuszek, Nicholas FitzGerald, Luke Zettlemoyer, Liefeng Bo, and Dieter Fox. A joint model of language and perception for grounded attribute learning. *arXiv preprint arXiv:1206.6423*, 2012.

[13] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.

[14] Marijn F Stollenga, Jonathan Masci, Faustino Gomez, and Jürgen Schmidhuber. Deep networks with internal selective attention through feedback connections. In *Advances in Neural Information Processing Systems*, pages 3545–3553, 2014.

[15] Stefanie Tellex, Pratiksha Thaker, Joshua Joseph, and Nicholas Roy. Learning perceptually grounded word meanings from unaligned parallel data. *Machine Learning*, 94(2):151–167, 2014.

[16] Jason Weston, Antoine Bordes, Sumit Chopra, and Tomas Mikolov. Towards ai-complete question answering: a set of prerequisite toy tasks. *arXiv preprint arXiv:1502.05698*, 2015.

[17] Jason Weston, Sumit Chopra, and Antoine Bordes. Memory networks. *arXiv preprint arXiv:1410.3916*, 2014.

[18] Kelvin Xu, Jimmy Ba, Ryan Kiros, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. *arXiv preprint arXiv:1502.03044*, 2015.

[19] C Lawrence Zitnick and Devi Parikh. Bringing semantics into focus using visual abstraction. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3009–3016. IEEE, 2013.

[20] C Lawrence Zitnick, Devi Parikh, and Lucy Vanderwende. Learning the visual interpretation of sentences. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1681–1688. IEEE, 2013.