VRNN 2018 - Predicting legal chess moves

Mateusz Kiebala
University of Warsaw
mk359758@students.mimuw.edu.pl

Tomasz Knopik
University of Warsaw
tk359778@students.mimuw.edu.pl

Janusz Marcinkiewicz
University of Warsaw
jm360338@students.mimuw.edu.pl

Abstract

One of the recent problems that gained a lot of attention in Deep Learning field was playing Go and Chess at the professional level. The models for these problems were assigning the probability of the win for certain moves and positions where the moves were generated deterministically. In our work we are checking if it is possible to learn a CNN network the rules of chess providing only image of the board before and after the move. We test two different models architectures: single CNN model which processes two input images as one concatenated; and Siamese CNN model which takes two images separately, extracts features combines them together predicting the result. These two approaches achieve 90% on training set and 85% on validation set.

1. Introduction

Recent achievements in Deep Learning allowed researchers to beat top players in Go and Chess using Deep Learning models. To learn model they have used techniques like Reinforcement Learning and Deep Convolutional Neural Networks to predict win percentages for given positions.

We believe that this problem is super important. If model can learn the rules of the chess, maybe it also can learn the rules of the world. Long standing question about single formula describing the world could be solved by CNN. It is still long way ahead of us but we believe we can make it and our work is first step towards this achievement.

2. Related work

We have struggled to find any papers or work done in this area. All the papers are targeting learning model to evaluate positions but not how to play the game. In end-to-end chess playing paper [1] authors discuss learning model to not only evaluate positions but also determining if the move is correct. We have used key insights from the paper to improve our models architectures. Another paper about predicting moves [2] provided a lot of useful information about using and not using different types of layers in our models.

3. Data

Chess is deterministic game and the moves as well as positions can be easily generated and encoded.

Data used in our experiments was generated using python chess library which allowed to save given position as images which then we converted into numpy arrays. We have created framework for generating random data with different outputs which were necessary for different types of models we have tested. The data was generated from random plays as well as positions and moves from book games. We have combined these two sources into datasets on which our models were tested. This way we wanted our model to prevent overfitting on specific types of plays. Illegal moves were created randomly but putting pieces in empty spaces on the chess board. The ratio of legal and illegal moves was around 50-50.

To train the models we have generated around 15 datasets with different random – book games ratio each of them containing about 5000 moves. In total we gained about 70000 moves to train on.

4. Methods

In our approach, we tried two different type of CNN models.

4.1. Single CNN model

One of the first models we train is CNN model which consists of couple convolution layers followed by the fully connected layers. The model, as an input, was feeded with

concatenated images: before and after move. In this model we wanted to check if convolution layers can extract more features having more information.

4.2. Siamese CNN model

Siamese CNN model consisted of two identical CNN networks which output was then concatenated and passed to the fully connected layers. This model was feeded with two inputs: image before and after move. Our understanding was that the model should independently extract features of each of the image and in the output have some kind of encoding which then was feeded to fully connected layers.

4.3. Deterministic encoding

In our last method we tried to see what accuracy can we achieve if we pass encoding of the chess board to fully connected layer without need of extracting features by convolution layers. This experiment was to see what is the bottleneck: too shallow representation of features extracted by convolution layers or fully connected layers. A term deterministic encoding means a chess board, encoded as {0,1} vector of length 768 (there are 12 different pieces which each of 64 positions can by occupied by).

5. Experiments

In the experiments we have tried different approaches for different types of models. In both types of models as well as deterministic encoding we achieved 85% on validation set.

Interestingly, the deterministic encoding had similar accuracy as the other models what suggest that convolution layers did a great job when extracting features and the bottleneck in the accuracy were fully connected layers. Worth noting is the fact, that thanks to the compact encoding representation, we were able to train a network on a much bigger dataset of roughly 200,000 moves. However we observed that it did not have signiticant impact on results, as final accuracy reached 87With this information we tried to implemented different set of fully connected layers but without further success.

5.1. Results by method

Method	Dataset size	No. epochs	Accuracy
Single	15,000	10k	50%
Siamese	9C	19C	50%
Single gray scale	10C	21C	87%

6. Conclusions

??

An analysis of the frobnicatable foo filter.

In this paper we present a performance analysis of the paper of Smith *et al*. [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.

[1] Smith, L and Jones, C. "The frobnicatable foo filter, a fundamental contribution to human knowledge". Nature 381(12), 1-213.

If you are making a submission to another conference at the same time, which covers similar or overlapping material, you may need to refer to that submission in order to explain the differences, just as you would if you had previously published related work. In such cases, include the anonymized parallel submission [**?**] as additional material and cite it as

[1] Authors. "The frobnicatable foo filter", F&G 2014 Submission ID 324, Supplied as additional material `fg324.pdf`.

Finally, you may feel you need to tell the reader that more details can be found elsewhere, and refer them to a technical report. For conference submissions, the paper must stand on its own, and not *require* the reviewer to go to a techreport for further details. Thus, you may say in the body of the paper "further details may be found in [**?**]". Then submit the techreport as additional material. Again, you may not assume the reviewers will read this material.

Sometimes your paper is about a problem which you tested using a tool which is widely known to be restricted to a single institution. For example, let's say it's 1969, you have solved a key problem on the Apollo lander, and you believe that the CVPR70 audience would like to hear about your solution. The work is a development of your celebrated 1968 paper entitled "Zero-g frobnication: How being the only people in the world with access to the Apollo lander source code makes us a wow at parties", by Zeus *et al*.

You can handle this paper like any other. Don't write "We show how to improve our previous work [Anonymous, 1968]. This time we tested the algorithm on a lunar lander [name of lander removed for blind review]". That would be silly, and would immediately identify the authors. Instead write the following:

We describe a system for zero-g frobnication. This system is new because it handles the following cases: A, B. Previous systems [Zeus et al. 1968] didn't handle case B properly. Ours handles it by including a foo term in the bar integral.

...

The proposed system was integrated with the Apollo lunar lander, and went all the way to the

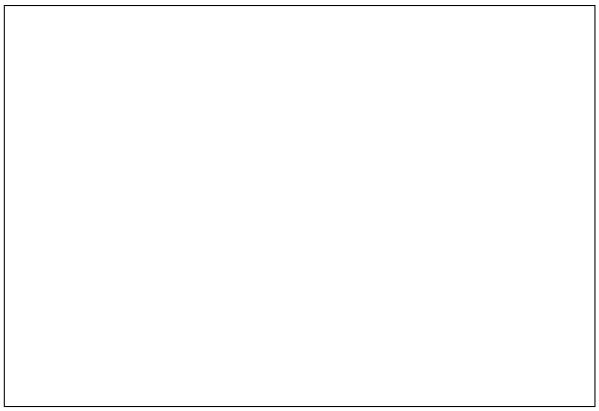

Figure 1. Example of caption. It is set in Roman so that mathematics (always set in Roman: $B \sin A = A \sin B$) may be included without an ugly clash.

> moon, don't you know. It displayed the following behaviours which show how well we solved cases A and B: ...

As you can see, the above text follows standard scientific convention, reads better than the first version, and does not explicitly name you as the authors. A reviewer might think it likely that the new paper was written by Zeus *et al*., but cannot make any decision based on that guess. He or she would have to be sure that no other authors could have been contracted to solve problem B.

FAQ

Q: Are acknowledgements OK?
A: No. Leave them for the final copy.

Q: How do I cite my results reported in open challenges?
A: To conform with the double blind review policy, you can report results of other challenge participants together with your results in your paper. For your results, however, you should not identify yourself and should not mention your participation in the challenge. Instead present your results referring to the method proposed in your paper and draw conclusions based on the experimental comparison to other results.

6.1. Miscellaneous

Compare the following:
`$conf_a$` $conf_a$
`$\mathit{conf}_a$` conf_a
See The TEXbook, p165.

The space after *e.g*., meaning "for example", should not be a sentence-ending space. So *e.g*. is correct, *e.g.* is not. The provided `\eg` macro takes care of this.

When citing a multi-author paper, you may save space by using "et alia", shortened to "*et al*." (not "*et. al.*" as "*et*" is a complete word.) However, use it only when there are three or more authors. Thus, the following is correct: " Frobnication has been trendy lately. It was introduced by Alpher [**?**], and subsequently developed by Alpher and Fotheringham-Smythe [**?**], and Alpher *et al*. [**?**]."

This is incorrect: "... subsequently developed by Alpher *et al*. [**?**] ..." because reference [**?**] has just two authors. If you use the `\etal` macro provided, then you need not worry about double periods when used at the end of a sentence as in Alpher *et al*.

For this citation style, keep multiple citations in numerical (not chronological) order, so prefer [**?**, **?**, **?**] to [**?**, **?**, **?**].

6.2. Type-style and fonts

Wherever Times is specified, Times Roman may also be used. If neither is available on your word processor, please use the font closest in appearance to Times to which you have access.

MAIN TITLE. Center the title 1-3/8 inches (3.49 cm) from the top edge of the first page. The title should be in Times 14-point, boldface type. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave two blank lines after the title.

AUTHOR NAME(s) and AFFILIATION(s) are to be centered beneath the title and printed in Times 12-point, non-boldface type. This information is to be followed by two blank lines.

The ABSTRACT and MAIN TEXT are to be in a two-column format.

MAIN TEXT. Type main text in 10-point Times, single-spaced. Do NOT use double-spacing. All paragraphs should be indented 1 pica (approx. 1/6 inch or 0.422 cm). Make sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

Figure and table captions should be 9-point Roman type as in Figures 1 and 2. Short captions should be centred.
Callouts should be 9-point Helvetica, non-boldface type. Initially capitalize only the first word of section titles and first-, second-, and third-order headings.

FIRST-ORDER HEADINGS. (For example, **1. Introduction**) should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one blank line after.

SECOND-ORDER HEADINGS. (For example, **1.1. Database elements**) should be Times 11-point boldface, initially capitalized, flush left, with one blank line before, and one after. If you require a third-order heading (we discourage it), use 10-point Times, boldface, initially capitalized, flush left, preceded by one blank line, followed by a period and your text on the same line.

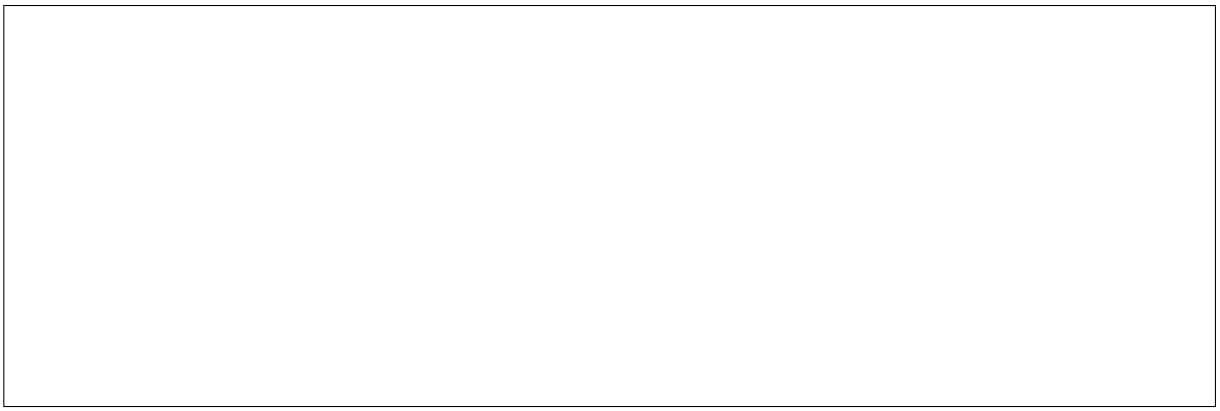

Figure 2. Example of a short caption, which should be centered.

Method	Frobnability
Theirs	Frumpy
Yours	Frobbly
Ours	Makes one's heart Frob

Table 1. Results. Ours is better.

6.3. References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [**?**]. Where appropriate, include the name(s) of editors of referenced books.

6.4. Illustrations, graphs, and photographs

All graphics should be centered. Please ensure that any point you wish to make is resolvable in a printed copy of the paper. Resize fonts in figures to match the font in the body text, and choose line widths which render effectively in print. Many readers (and reviewers), even of an electronic copy, will choose to print your paper in order to read it. You cannot insist that they do otherwise, and therefore must not assume that they can zoom in to see tiny details on a graphic.

When placing figures in LaTeX, it's almost always best to use `\includegraphics`, and to specify the figure width as a multiple of the line width as in the example below

```
\usepackage[dvips]{graphicx} ...
   \includegraphics[width=0.8\linewidth]
                   {myfile.eps}
```

References

[1] E. David, N. S. Netanyahu, and L. Wolf. Deepchess: End-to-end deep neural network for automatic learning in chess. *CoRR*, abs/1711.09667, 2017.

[2] B. Oshri. Predicting moves in chess using convolutional neural networks. 2015.