Who is actually winning ?

Authors: Gautam Shende, Pradeep Aitha, Sujit Singh

CSE 712: Human Decision Making and Machine Learning

Instructor: Dr. Kenneth Regan

Date: 05/19/2018

OVERVIEW

1. Hypothesis
2. Model and Methodology
3. Biases and limitations
4. Results & Inferences
5. Final Conclusion
6. Future Scope
7. References

TM

1) Hypothesis

Hypothesis

1. Stockfish evaluation alone can not determine “who is winning”.

2. A more comprehensive approach is needed for predicting an ‘Who is winning ?’. This requires to consider a lot of other factors.

3. Example :
   Kasparov(Opening) vs Karpov (Endgame)
   - Kasparov having early advantage does not signify winning
   - Karpov having early advantage does signify a better chance

2) Model and Methodology

Model

1. Features:
   Input is details about a move
   - ELO: ELO_White, ELO_Black
   - Turn: Turn, Move#
   - Positional arguments:
     QueenWhite, QueenBlack, RookWhite, RookBlack, BishopWhite, BishopBlack, KnightWhite, KnightBlack

   - Evaluations:
     Stockfish Bestmove_Eval

   - General Info:
     PlayerNames, Year, Type of tournament

Methodology

1. DataSet:
- AA_..._.txt files from metallica
- ~74 Files
- ~7 Million Moves
- ~1 Million Moves used for training Large Model
- ~100,000 Moves used for training Small Model (Flexibility)

2. Parsing:
- Regex and lambda for processing input from files
- Filtering:
    - Drop Cities
    - Tournament filtering (U,Men,Women,General)
    - removing empty entries in ELO (unr) and Year (?)
    - Player name hashing
    - Parse FEN

Methodology

3) Major steps:

- Ran Random Forest Classifier and Regressor after selecting 1 Million Random Moves from the original 7 Million data set.

- Classifier was used in both Models(Large and Small) and Regressor was used only in the smaller model.

- Testing of how exclusion and inclusion of different parameters affects prediction in the model.

- Testing whether using only stockfish evaluation is good enough ?

Methodology

3) Major steps(contd):

(Work in Progress)

- Building a front end:
- Model simulation (Output of a given input)
- Live Game Analysis

TM

3) Biases and Limitations

Biases and Limitations

1) Limited ELO differences. Model wouldn’t be too appropriate for high ELO difference

2) Feature variance. Example: More data in fixed range of ELO (~2200 to 2600 range)

3) Limited Data Set (~ 1 million in Large model and ~100,000 in Smaller)

4) Games fixed around recent years. More data from recent years. (Which is not that bad

5) Hashing names limited to hashing function (Time vs accuracy tradeoff). Use other method for identifying person (Birth date + name hasing for example is more accurate)

4) RESULTS & INFERENCES

Random Forest on Large Dataset

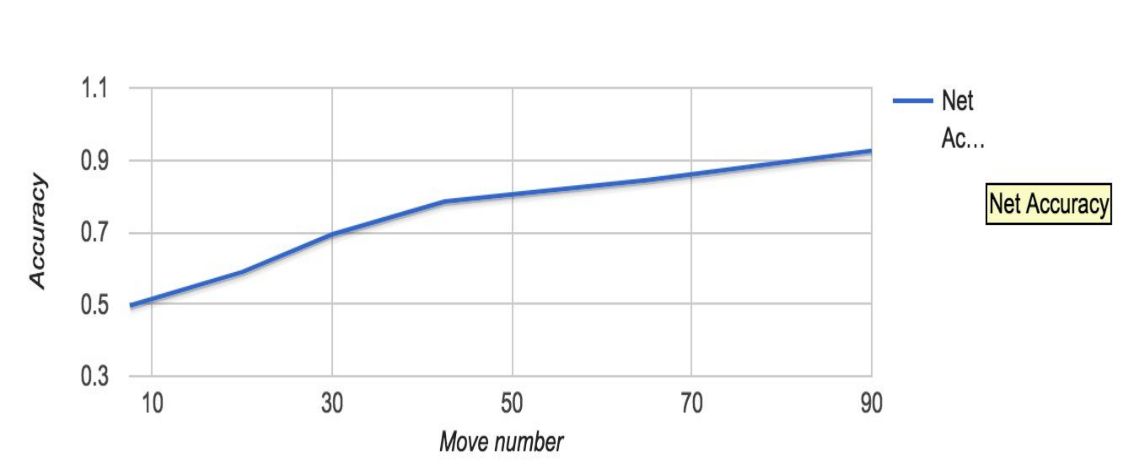
Random Forest Classification of games on testing set for large data set ~ (100,000)
Accuracy: 0.6566

	Classified B	Classified D	Classified W
Black Wins	16373	7243	3740
Draw	4554	25989	6991
White Wins	2827	8984	23298

Immediate Observations:

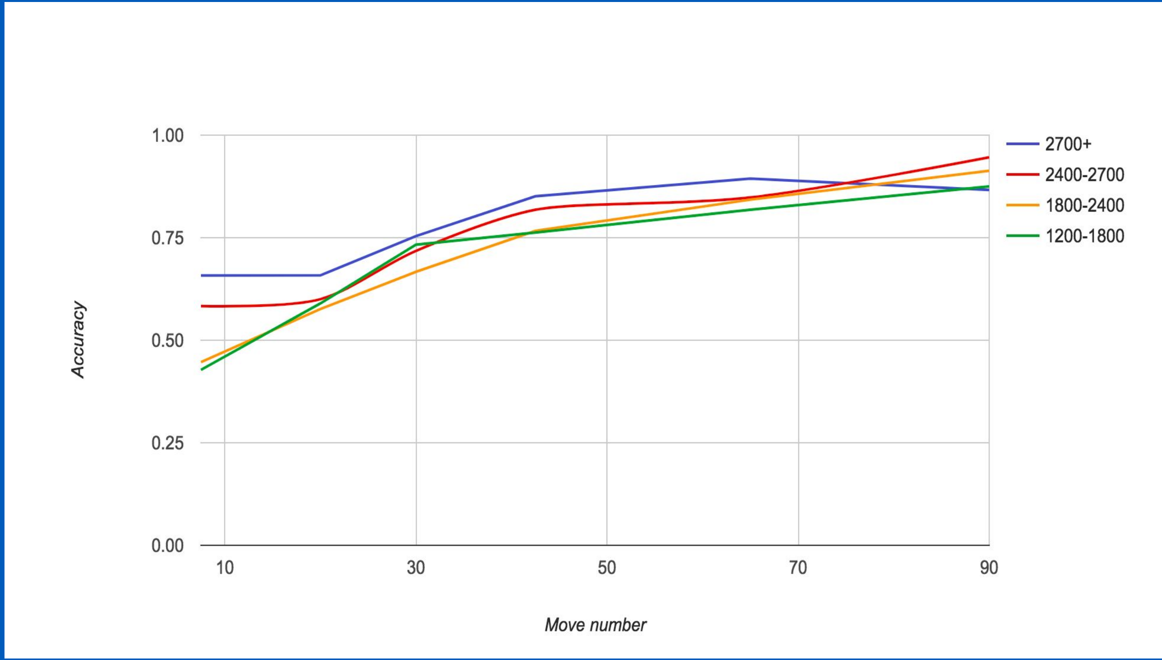
- White wins are more, draws are max and Black wins are least
- Max misclassifications are for draw (~16k), followed by White(~11k) and Black(~7k).

a) Accuracy of large set with Move



- Accuracy increases with increase in number of moves as expected (Very hard to predict who is going to win in initial stages).

b) Accuracy vs Move # for different ratings



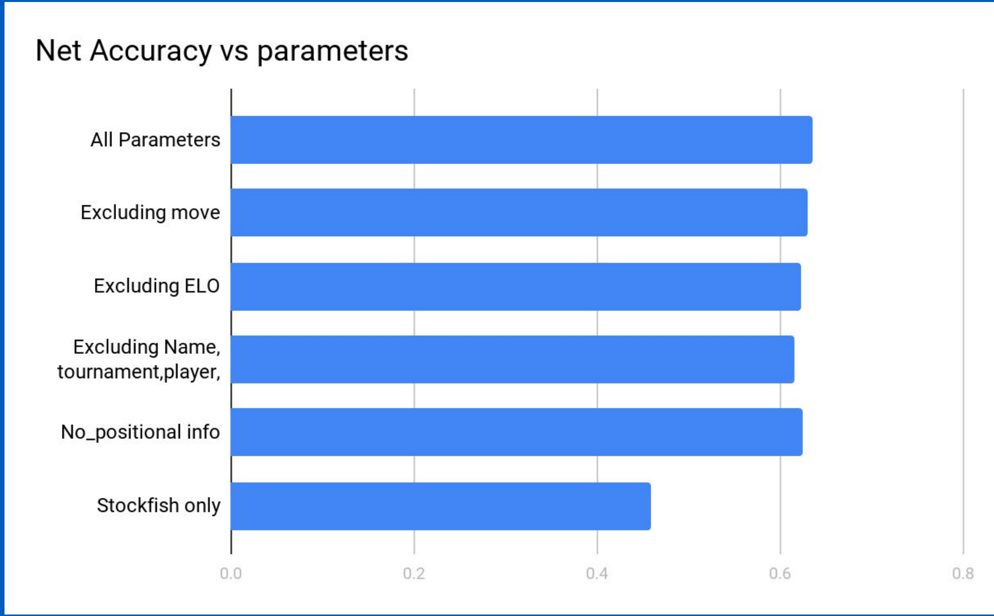
b) Accuracy vs Move# Inferences

Inferences:

- Accuracy is maximum for Highest ELO ratings (2700+) until the endgame.
- Accuracy of GrandMaster range (~2400-2700) starts at second and ends up at the top towards the end game
- The ranges between (1200-1800) and (1800-2400) keep fluctuating
- Important inference is that it is easier to predict outcomes as we go higher in the ratings

TM

c) Accuracy vs Move # for different ratings (smaller data)



-

Params	All params	move’	ELO’	(name,tourn,year)’	Positional info’	Stockfish only
Accuracy	0.636	0.63	0.623	0.616	0.625	0.458

b) Accuracy vs Move # for different ratings

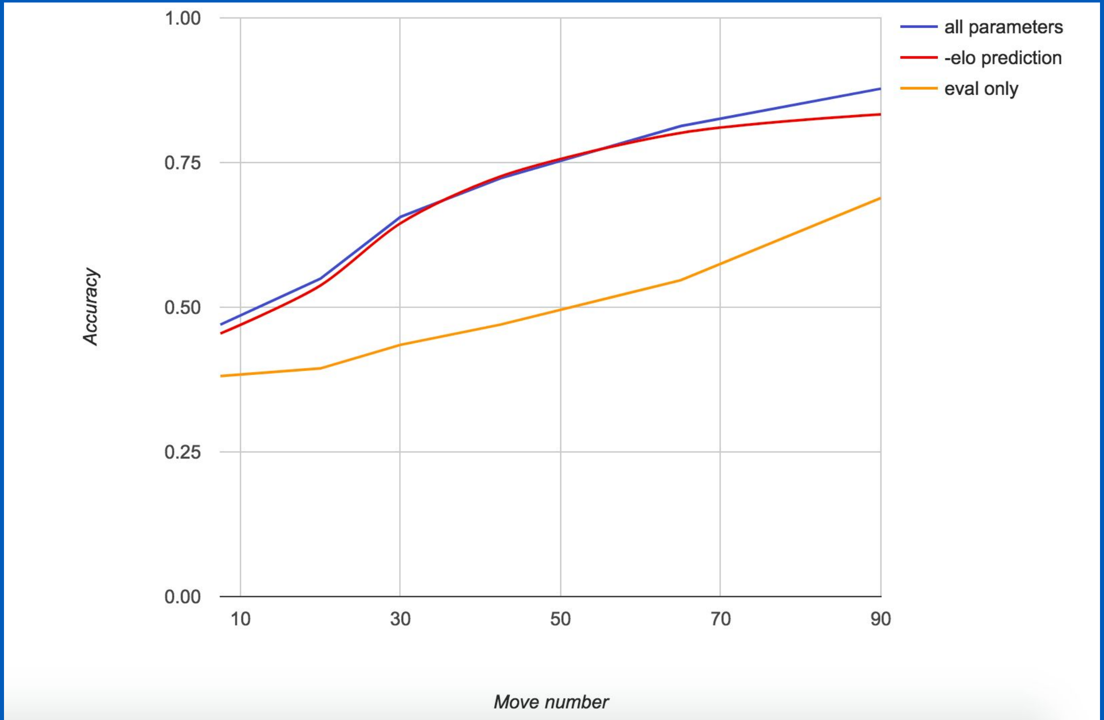
Inferences:

- Excluding name tournament and year which we thought wouldn’t have much impact on accuracy has actually the most impact

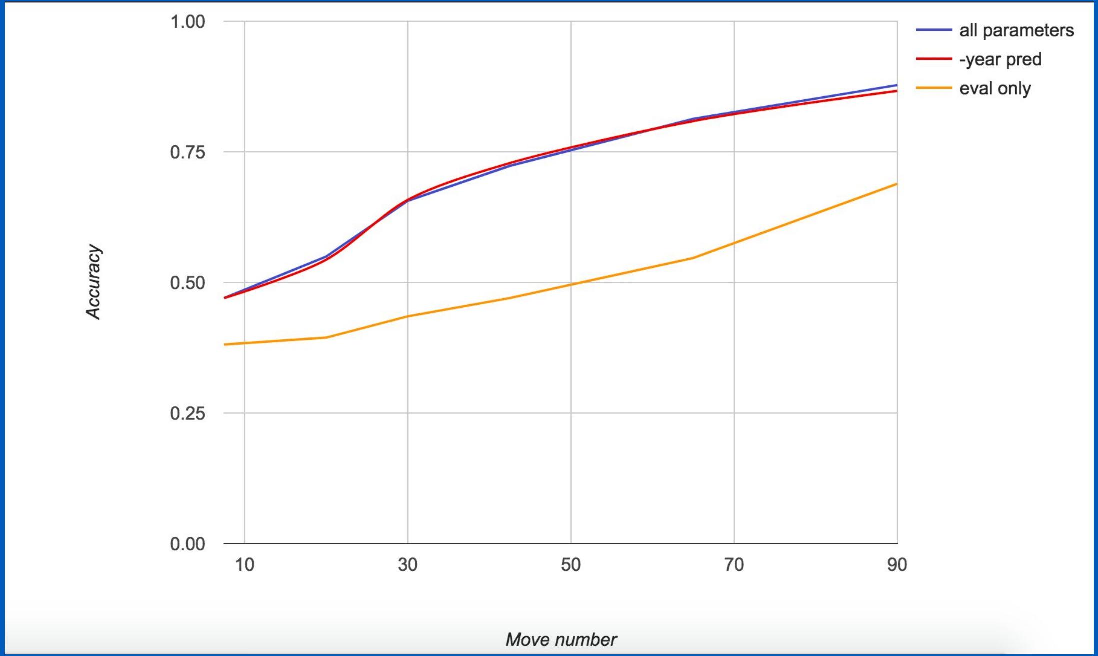
- Excluding ELO was second most impactful and strongest when we consider individual features (as expected)

- Excluding move number and Move position did result in slight drop in accuracies but those were even more insignificant

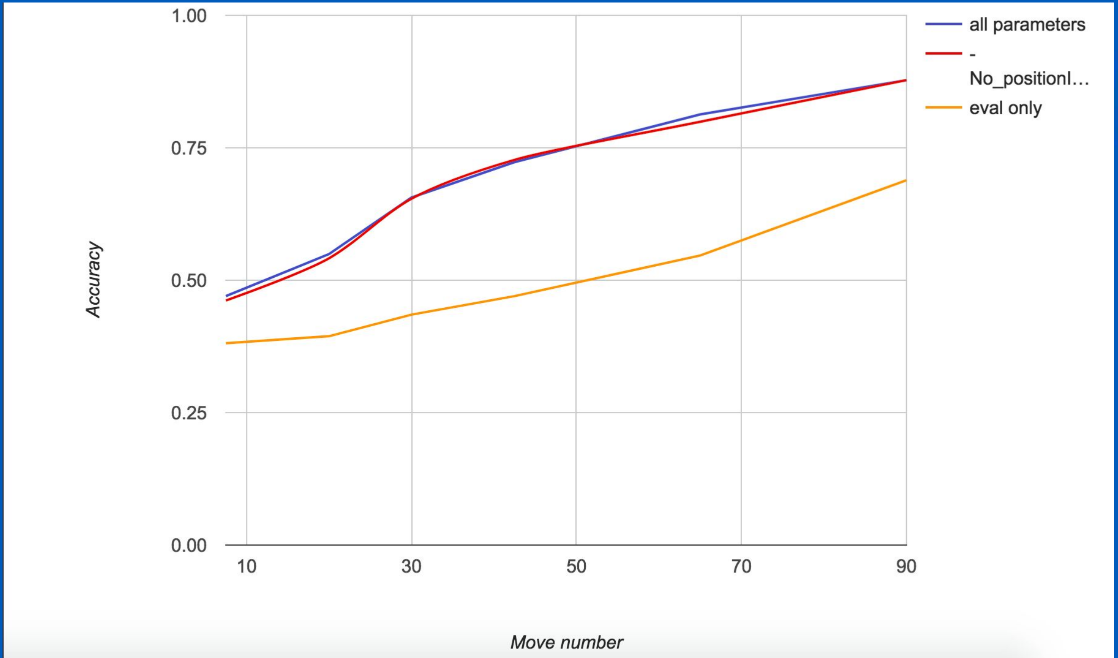
d) Dropping ELO



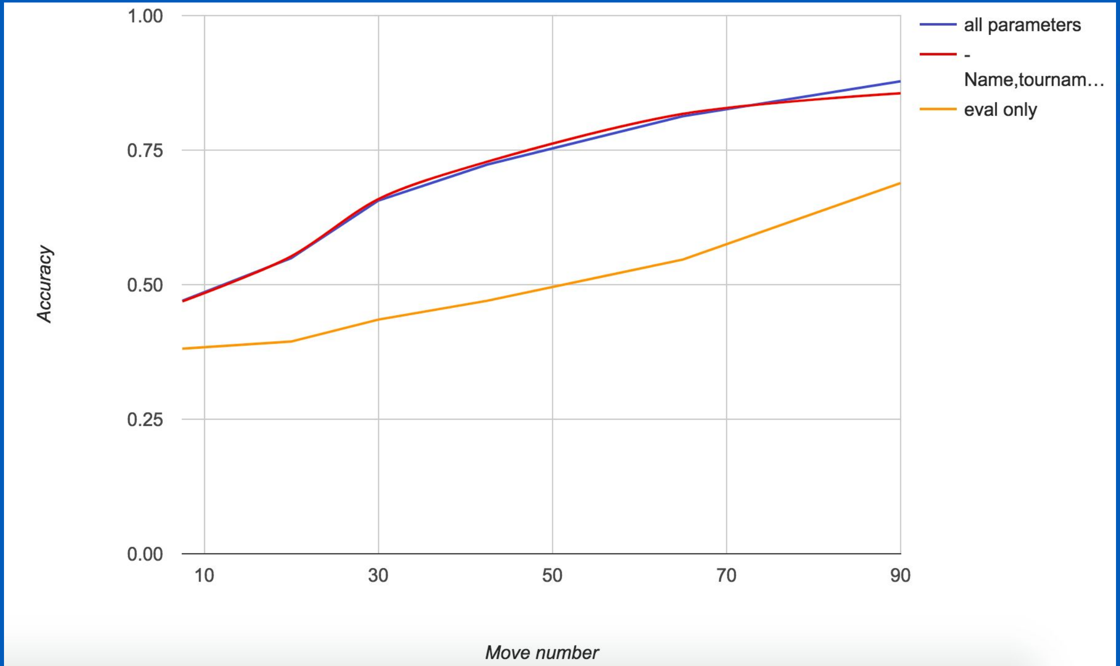
e) Dropping year



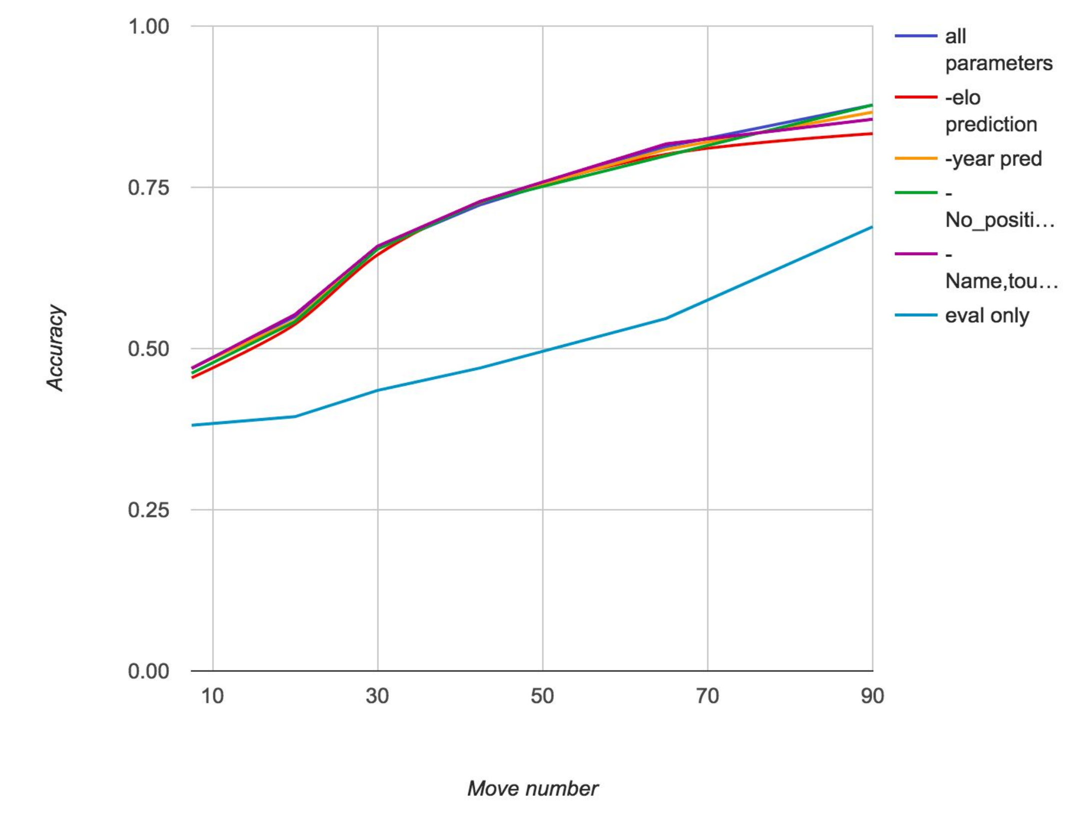
f) Accuracy vs Move # for different ratings



g) Dropping Name, Tournament, Year



h) Everything one one graph (A mess!)



i) Dropping parameters inferences

Inferences:

- Dropping one or two parameters doesn’t affect the model much

- Only slightly significant observation is that in really higher ELOs (2700), knowing the ELO doesn't help in predicting outcome towards the endgame.

- Dropping all parameters does have significant impact on the accuracy

5) Final conclusion

Conclusion

1) Using just stockfish evaluation for predicting results is surely one of the best features to consider

2) However, considering other features certainly helps in yielding a more comprehensive

3) Practical ?
   - Not that much!
   - Maybe 2 or 3 features can be considered however increasing parameters more than that doesn't yield a significant increase in performance in comparison to the increase in running time.

6) FUTURE SCOPE

FUTURE SCOPE

1) Including Player type (maybe from some other group_ as a parameter

2) Do it for specific 20 players so model can analyse their game in detail. How do parameters now affect the accuracy ?

3) Using AWS for running 7 million data file

TM

7) REFERENCES

REFERENCES

1) Moves Dataset - Dr. Kenneth Regan (metallica.cse.buffalo.edu)

2) Scikit-learn.org

3) https://machinelearningmastery.com/save-load-machine-learning-models-python-scikit-learn/

4) Chess24.com (Scraping)