CSCE 489 - Introduction to Data Science
Final Project Proposal

Juan Burgos, Brady Pacha

March 29, 2013

Project Goals

Our goal is to design and implement a platform by which chess strategies can be analyzed in an in-depth and visual manner. We will provide an animated heat-map based using player turns as the "time" unit to show the probability distribution of specific chess pieces placements on the board and how they change over time. We'll aggregate location of specific pieces, over all of the games that we have on record, and will provide the probability of a piece existing at any board location given the current turn in the game.

Imagine a turn-by-turn animation that shows the probability of, for example a pawn, as the probability of it existing at any legal point on the board changes. If you want to know the location of some pawn after say 20 turns, simply pause the animation at the 20th turn. We will then provide that pawn's location probability distribution, visualized as a heat-map, over the entire board by analzying that pawn's location at the 20th turn for all 50,000 games.

Extra Goals - If we have ample time...

Given a heat-map for any piece at any given turn, if we have time, we would like to try and provide survivability data for this piece, player win/loss predictions given any game state where the piece is at the current location given the current turn, average remaining game duration given the current board state(s), and/or a recommendation of the next best move.

Each of these items would be very interesting to investigate and possibly deliver to provide the user with additional information. Recall that these are items we'd like to explore *if* we have enough time.

Project Motivation

This application can be a useful educational tool for students learning to play chess. They can see what common piece placement strategies have been used in the past, including the most common locations of pieces after x amount of turns and possibly some statistics associated with those piece placements. Chess enthusiasts may find an interest in the applicatino because it would provide a much deeper insight into strengths/weakness of strategies and relationships between different strategies.

Data

0.1 Data Format

We have found several online repositories containing the list of all moves, including game meta-data, for roughly 50,000 chess games, many of which are master level matches. All of our data comes in PGN format, an example of which is provided below.

```
1   [Event "LAT-ch"]
2   [Site "LAT"]
3   [Date "1952.??.??"]
4   [Round "?"]
5   [White "Klasup, Karlis"]
6   [Black "Tal, Mihail"]
7   [Result "0-1"]
8   [WhiteElo ""]
9   [BlackElo ""]
10  [ECO "D46"]
11
12  1.d4 Nf6 2.c4 e6 3.Nf3 d5 4.Nc3 c6 5.e3 Nbd7 6.Bd3 Bb4 7.a3 Ba5 8.O-O O-O
13  9.Ne5 Nxe5 10.dxe5 dxc4 11.Bxc4 Nd7 12.f4 Qe7 13.b4 Bb6 14.Qb3 f6 15.Bxe6+ Kh8
14  16.Ne4 fxe5 17.Kh1 exf4 18.exf4 Nf6 19.Ng5 Ne4 20.Nf7+ Rxf7 21.Bxf7 Nf2+
15  22.Kg1 Nd3+ 23.Kh1 Qe2 24.Bb2 Bh3 25.Bxg7+ Kxg7 26.Bd5 cxd5 27.Qxd5 Kh8 28.Qg5 Bd4
16  29.gxh3 Rg8 30.Rae1 Nxe1   0-1
```

0.2 Example Source

The following link is to one of the repositories that we visited: world-chess-championship-1886-2012. The package contains info on all games from the 1886 - 2012 World Chess Championship tournaments in PGN format.

1 Current Design Ideas

1.1 Project Tools

We may consider making use of the python parser library called pgnparser for extracting the useful information.

1.2 Implementation

We've considered using D3 to produce the required visualizations for our application. We are still looking into how this may be done. We have already found matplotlib visualizations of chess games, but would like to provide this service as an online application. This would be the frontend portion of the project. In the backend, we will use a database, possibly mongodb, to store the data required to produce an animated heat-map.

The data needed to produce the heat-map will be processed via AWS by analyzing all the chess games to capture piece location statistics and histories.