Football-o-Genetics

(version 2013/08/25)

Michael A. Alcorn

Installation

1. Install a Java Runtime Environment (JRE) on your computer.

    i. To check if a JRE is already installed, run:

    ```
    java -version
    ```

    from the terminal console (Mac, Unix-like) or the command prompt (Windows). If you see:

    ```
    java version "1.6.0_38"
    Java(TM) SE Runtime Environment (build 1.6.0_38-b05)
    Java HotSpot(TM) 64-Bit Server VM (build 20.13-b02, mixed mode)
    ```

    or something similar as your output, move on to step 2.

    ii. To install a JRE on a Linux machine, I recommend following the directions found on this webpage. To install a JRE on a Windows machine, download and run the executable file available on Java's website.

2. `cd` to the directory containing the "Football-o-Genetics.jar" file.

3. Run:

    ```
    java -jar Football-o-Genetics.jar
    ```

    Note: on Windows, double-clicking "Football-o-Genetics.jar" will start the application.

Collecting Play-By-Play Data

(Using the 2012 Carolina Panthers as an example)

1. Create a folder named "Data".

2. In the newly created "Data" folder, create a folder named "Games".

3. Go to www.espn.com.

4. Click on the "NFL" header.

5. Select "Carolina Panthers" from the "Teams" pull-down menu.

6. For each game on the schedule:

    i. Click on the score under the "RESULT/TIME (ET)" header.
    ii. Click on the "Play-By-Play" tab.
    iii. Click on "All".
    iv. Copy all of the text spanning from "1st Quarter Play by Play" to the final "DRIVE TOTALS..." and paste it into a plain-text text editor (e.g. Gedit or Notepad).
    v. Save the file in the "Games" folder.

(a) **Note: the name of the file MUST be identical to the name that is assigned to the opponent in the play-by-play descriptions.** For example, the play-by-play file for the Panthers vs. Buccaneers game should be named “Tampa Bay Buccaneers”. Make sure the file name does **NOT** include a file extension (e.g. “.txt”).

(b) **Note: if an opponent occurs more than once in the target team's schedule, the opponent's name should be replaced with “Opponent's Name #” in the play-by-play text (with # referring to the number of meetings the two teams have had thus far).** For example, in the second Panthers vs. Buccaneers game, each occurrence of “Tampa Bay Buccaneers” should be replaced with “Tampa Bay Buccaneers 2” in the play-by-play text, and the file should be saved as “Tampa Bay Buccaneers 2”.

1. **Tip: use the find and replace function found in most text editors to quickly replace team names in play-by-play text.** The find and replace function can typically be brought up by pressing Ctrl-H on the keyboard (i.e. holding the Control key down and then pressing the H key).

Using the Application

1. “**Target Team**” **–** indicates the team that is the subject of the optimization. The contents of this field MUST match the name used by the target team in the play-by-play descriptions. In the case of the Carolina Panthers, the “**Target Team**” would be “Carolina Panthers”.

2. “**Interceptions**” **–** opens a .csv file containing the interception rates for the quarterbacks who will be included in the optimization.

3. “**Fumbles**” **–** opens a .csv file containing the fumble rates for the running backs who will be included in the optimization.

******************************Interceptions/Fumbles Formatting Guidelines******************************

The interceptions/fumbles file **MUST** be properly formatted for the application to run smoothly. I have provided some guidelines below, and I have also included some example data files (“InterceptionRates.csv” and “FumbleRates.csv”) in the “ExampleData.zip” package for reference.

i. The interceptions/fumbles file should have one line for each quarterback/running back.
ii. Each row should include the player's name, followed by a comma (“,”), followed by the player's interception/fumble rate. **Note: there are no spaces between the end of the player's name and the comma, nor between the comma and the start of the player's interception/fumble rate.**
iii. The interception/fumble rate for a given player is calculated by dividing the number of interceptions/fumbles a quarterback/running back had during a season by the number of passing/rushing attempts the quarterback/running back had during the season.

******************************Interceptions/Fumbles Formatting Guidelines******************************

4. “**Games Folder**” **–** opens a directory containing the play-by-play files. Simply select one of the game files contained in the “Games” directory.

5. “**Raw Data**” – if you already have a **properly formatted** .csv file containing play-by-play data, then this button can be used in place of the “**Games Folder**” button. The “RawData.csv” file included in the “ExampleData.zip” package is an example of a properly formatted raw data file. The raw data file should have ten columns, with each row representing a different play from the season. The first column should contain the opponent's name, the second column should contain the quarter, the third column should contain the down, the fourth column should contain the distance, the fifth column should contain the side of the field, the sixth column should contain the yard line, the seventh column should contain the player's first initial, the eighth column should

contain the player's last name, the ninth column should indicate whether the play was a rush or a pass (**Note: sacks and scrambles should be considered passes**), and the tenth column should contain the yards gained on the play.

6. "**Get Stats**"

   i. Reads in the play-by-play files.
   ii. Tidies the play-by-play files and places these new files into a "Cleaned Games" folder.
   iii. Moves the raw play-by-play files into the "Raw Games" folder.
   iv. Creates a .csv file named "RawData.csv" and places it in the "Results" folder. The file contains the opponent, quarter, down, distance, field position, player who executed the play, play ("Pass" or "Rush"), and yards gained on the play for every play from the target team's season.

7. Input fields

   i. "**QB A**", "**QB B**", "**RB A**", and "**RB B**" **–** the players that will be included in the optimization. The pull-down boxes will be auto-populated once you click the "**Get Stats**" button. **Note: "None" can be selected for "QB B" and/or "RB B".**
   ii. "**Drives**" **–** the number of drives that will be used to estimate each individual's fitness. More drives can provide a better estimate of fitness, but will increase the time necessary to complete the optimization.
   iii. "**Generations**" **–** the number of evolutionary steps that the population will take in the optimization. More generations can result in better solutions, but will increase the time necessary to complete the optimization.
   iv. "**Population Size**" **–** the size of the population that will be used to evolve a play calling strategy. A larger population size can result in better solutions, but will increase the time necessary to complete the optimization.
   v. "**Use Distance?**" **–** determines whether player stats should be binned according to down and distance (e.g. "3rd and Short" or "2nd and Long") or down only (e.g. "1st Down" or "2nd Down").

8. "**Optimize**" – evolves a solution and outputs the results to a .csv file named "OptimizationResults.csv", which can be found in the "Results" folder. To monitor the progress of the optimization, run the application from the terminal or command prompt.

9. "**Start Over**" – resets the app. **Note: the "Start Over" button must be clicked before each new optimization.**

Interpreting Your Results

1. "**Trial**" – indicates whether the play calling strategy was observed or evolved. The first row will always contain "Observed (N)", with "N" representing the number of offensive drives conducted by the target team during the season.

2. "**Fitness**" – $\frac{\#\ simulated\ drives\ resulting\ \text{in}\ a\ touchdown}{\#\ simulated\ drives}$

3. "**QB A**", "**QB B**", "**RB A**", and "**RB B**" – the players used in the play calling strategies.

4. "**1stPass**" through "**3rdShortRush**" – the probability with which "**QB A**" was used as the quarterback on a pass play or with which "**RB A**" was used as the running back on a rush play in the given situation. "**Short**" implies a distance of two yards or less, "**Medium**" implies a distance of four to six yards, and "**Long**" implies a distance of seven yards or more.

5. "**1st**" through "**3rdShort**" – the probability with which a rush play was called in the given situation.