

CSC2541: Sports Analytics (Challenge 2)

February 25, 2017

1 Dataset

You are provided with Sportvue tracking and play-by-play data for 42 Toronto Raptors games played in Fall 2015. The data is collected by multiple cameras installed around the court which detect and track the players. The cameras operate at 25 frames per second and cover the whole court throughout the game. For each frame the position of each player and the ball is provided. The game is split into particular events and annotated with playbyplay data. You can download the data by clicking [here](#) (403 mb).

The data is structured as follows:

- After unzipping the data file using <http://www.7-zip.org/>, you will find all the Raptors games in `raptors/data/`.
- Each game is saved as a python pickle file. Take for example `0021500463.pkl` which you can load by:

```
import pickle
game_name = '0021500463.pkl'
game = pickle.load(open('CHANGE_THIS/raptors/'+game_name, "rb"))
```

- `game` is a python dictionary object with the following keys:

```
game.keys()
['gamedate', 'gameid', 'events']
```

- `game['gamedate']` and `game['gameid']` contain respectively the game date and the id assigned by Sportvu to the game:

```
game['gamedate']
u'2015-12-28'
game['gameid']
u'0021500463'
```

- If you append the gameid to the end of `http://stats.nba.com/game/#!/`, for example `http://stats.nba.com/game/#!/0021500463/`, then you can view all the data except the tracking data on NBA's website. Perhaps this can be useful for further data mining.
- All the important tracking and play by play data as well as team descriptions are contained in `game['events']` which is python list. The length of each list is roughly equal to the number of different offensive events happening in the game. For example, the current game has this many events:

```
len(game['events'])
231
```

- Note that each game has a different number of events.
- Each event is a dictionary. For example:

```
event = game['events'][100] # take the 100th event
event.keys()
```

```
['playbyplay',
 'moments',
 'visitor',
 'end_time_left',
 'start_time_left',
 'home',
 'quarter',
 'orig_events']
```

- `event['home']` and `event['visitor']` are dictionaries with info about the teams:

```
event['home'].keys()
[u'abbreviation', u'players', u'teamid', u'name']
```

- `event['quarter']` specifies the period of the game.
- `event['start_time_left']` and `event['end_time_left']` specify the time left in seconds to the end of the period at the beginning and at the end of the event respectively:

```
event['start_time_left']
149.18
event['end_time_left']
126.23
```

- `event['moments']` is a list with all the player and tracking data. Consecutive elements of the list correspond to consecutive frames of the event. Note that each 25 consecutive elements of the list correspond to one second of the game. For this particular event, there are 575 moments corresponding to 23 seconds of the game:

```
len(moments)
575
```

- As an example, lets look at the 200th moment:

```
moment = moments[200]
```

- `moment` is a list of lists.

```
moment
```

```
[2,
```

```

1451354059013,
141.18,
23.38,
None,
[[-1, -1, 18.93906, 38.41749, 3.86873],
[1610612741, 2200, 4.62944, 25.44202, 0.0],
[1610612741, 201166, 60.48311, 44.86698, 0.0],
[1610612741, 201959, 34.09068, 15.53038, 0.0],
[1610612741, 201565, 17.29725, 36.95754, 0.0],
[1610612741, 202710, 54.72189, 4.32611, 0.0],
[1610612761, 2449, 44.61876, 15.76737, 0.0],
[1610612761, 201960, 53.06096, 39.65171, 0.0],
[1610612761, 200768, 40.51115, 25.8428, 0.0],
[1610612761, 201942, 63.45327, 13.47447, 0.0],
[1610612761, 202685, 43.60377, 27.70004, 0.0]]]

```

- `moment[0]` is the period of the game.
- `moment[1]` is the timestamp in milliseconds of the moment/frame.
- `moment[2]` is the time in seconds left to the end of the period.
- `moment[3]` is the time left to the end of the shot clock.
- `moment[5]` is the tracking data.
- `moment[5][0]` is the ball position data where

```

moment[5][0]
[-1, -1, 18.93906, 38.41749, 3.86873]

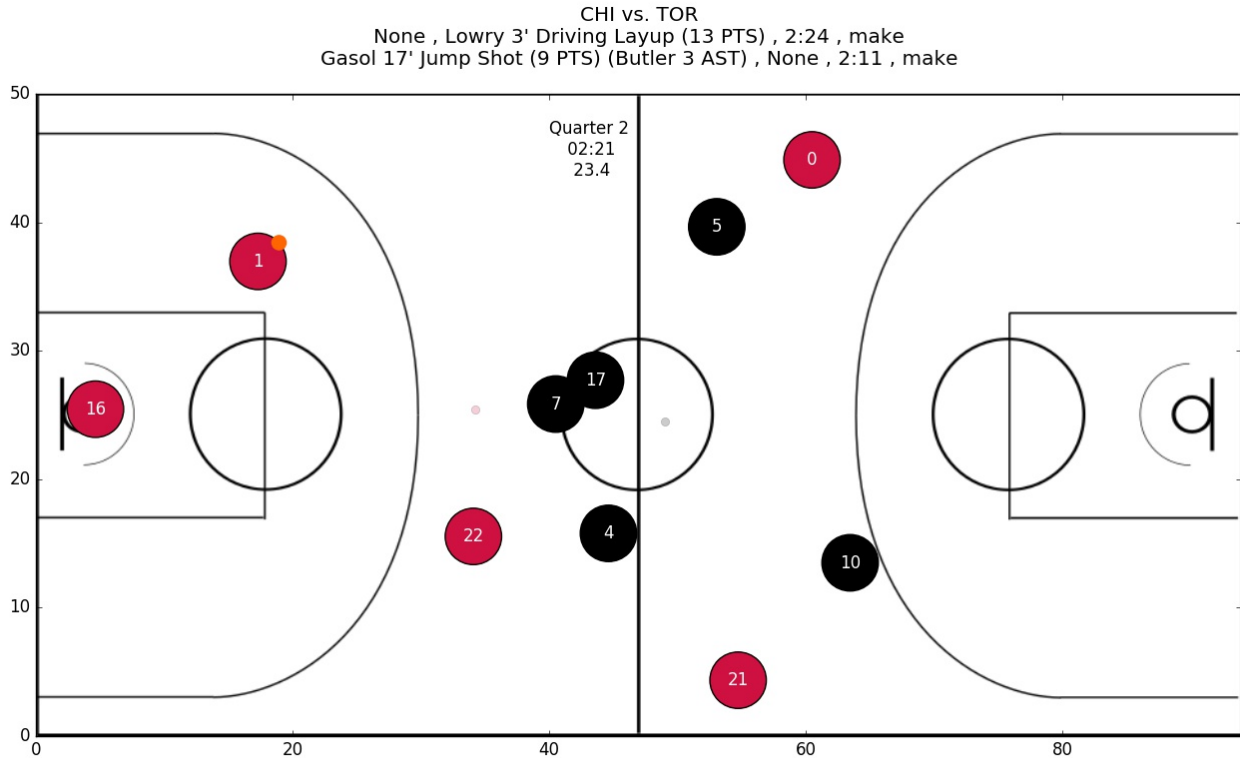
```
- The last three elements of the list correspond to (x, y, z) coordinates of the ball.
- `moment[5][1:]` are the player position data.
- For each player, say `moment[5][2]`, the first element of the list is the team id, the second is the player id, the third and fourth are the (x, y) position of the player. You can ignore the last element:

```

moment[5][2]
[1610612741, 201166, 60.48311, 44.86698, 0.0]

```
- Finally `event['playbyplay']` is a pandas dataframe that contains the play by play data of the event.
- We have provided you with starter code to load and visualize the data. It's located in `raptors/code/`. You have to set the right directories and make sure that you have installed all the required python packages. Also you can click on this [URL](#) to get more ideas on the data visualization.

- Here is the visualization of the moment we used in our example:



2 Challenge

The goal of this project is to be able to learn in an unsupervised fashion activities, events, plays, etc that happen in the game. You can use any code available on the web so long as you properly acknowledge it. The limit is your imagination.

Important Note: If the amount of data is too cumbersome for your computational resources, you can reduce the size of your training set at NO penalty.

3 Submission

A report (paper format) should be submitted explaining what you have done and relating it to the state-of-the-art. A presentation will be given in class as well.