NFL_BDB_feedback_response_Brad_Nott

April 11, 2019

```
In [1]: import os
        import pandas as pd
```

```
In [2]: os.chdir('C:/Users/Brad/Desktop/NFL Data/Big-Data-Bowl-master/Big-Data-Bowl-master/Data
```

For a given play, is the `first_contact` tag unique to a particular player?

- Ideally it should be unique to the ball carrier to support measuring rusher yards gained after contact

```
In [5]: # Load a sample tracking data file
        df_test = pd.read_csv('tracking_gameId_2017090700.csv')

        # Subset to only keep the rows with a first_contact tag
        df_contact = df_test[df_test['event'] == 'first_contact']

        # check how many players on a given play have the first_contact tag
        df_contact.groupby(['playId'])['event'].value_counts().unique()
```

```
Out[5]: array([22,  1, 21], dtype=int64)
```

In this particular game, across all plays there are only 3 unique-sized groups of players who have the `first_contact` tag: 22, 21, or 1 players. In other words, it is common for each player involved on a play to have this tag.

With that in mind, let's also check if all the *times* associated with those tags are identical. If they are, then the `first_contact` tag itself is not necessarily unique to when a particular player makes contact with the defense.

```
In [8]: # Are the times the same?
        df_contact.groupby(['playId'])['time'].value_counts().unique()
```

```
Out[8]: array([22,  1, 21], dtype=int64)
```

So all players on a given play are assigned the `first_contact` tag at the same exact times. So what does the `first_contact` tag even mean?

If we cannot resolve the context of the tag using a player name or a time, we will try to derive its meaning relative to other tags.

To establish the proper context, we need to know what event tags commonly occur immediately before the `first_contact` tag.

```
In [16]: # Prepare to store all tags that immediately precede a first_contact event tag
         precedes_first_contact = []

         # Begin with no previous event tag
         previous = None

         # For a given tracking data file, loop over event column
         for event in df_test['event']:

             # Assign a new current tag
             current = event

             if type(event) != str:

                 # Ignore nan tag; continue loop
                 continue

             elif current == 'first_contact':

                 # Store event tag that came before a first_contact event tag
                 if previous not in precedes_first_contact:
                     precedes_first_contact.append(previous)

             # Assign current tag as the new previous tag; advance the loop
             previous = current
```

While each player has the `first_contact` tag at the same timestamp, the tag only ever appears following these tags:

```
In [17]: precedes_first_contact

Out[17]: ['kick_received',
          'pass_outcome_caught',
          'handoff',
          'fumble',
          'run',
          'ball_snap',
          'lateral',
          'pass_arrived',
          'punt_received']
```

This list of event tags seems to suggest that the `first_contact` tag is supposed to describe the moment that a **ball carrier** makes contact with a defender.

Conclusion: the `first_contact` tag is not unique to a particular player, but it does appear to be unique, in context, to a ball carrier. Therefore, if we know who the ball carrier is, we can use the `first_contact` tag to explain when that ball carrier makes contact with a defender. Then, to assess yards gained after contact, all we need to do is measure distance moved downfield from the `first_contact` tag until the time when the player is down, scores a touchdown, runs out of bounds, etc.