Discussion: Creative analytics for UIUC Football Team

Input

Data from Kingsley/football. This is all about the defensive side of the ball for now but offense is on the horizon. I told him that getting this ahead of a meeting would be helpful so that we can have at least looked at it before he explains it.

Kingsley is doing a good job of dreaming out, including about virtual reality – which he used while with the Tampa Bay Buccaneers, and machine learning. He does want to keep output from all of this data simple per Lovie's desires for simplicity.

Data Available

For those who aren't familiar with football, games are split into plays. Teams alternate offense and defense, and have different players for each side. There are several players for each position who will rotate in and out of the game between plays based on the situation (i.e. your big slow guy might come in if you only need 1 yard to score, but a smaller faster guy if you need a long run). So only 11 people on the field per team at once, but ~52 players on a team roster in total.

You have 4 downs (plays) to gain 10-yard increments and if you can't advance 10 yards in those 4 plays, you turn the ball over to the other team's offense and they try to go the other direction. If you do advance, you get "first down" and 4 more chances to get 10 MORE yards... repeat until you either score (reach the endzone at one of the of the field) or you turn the ball over.

The defensive data that was shared can be split into a few groups:

- **Single Game Participation and Positioning** (x4 in sample)
  - Data on one game between two teams.
  - Each row describes one player, during one play
  - For each player
    - on_offense - on offense or defense? will be the same for everyone on the same team for a given play
    - *defined in CFB_Participation_Decode below:*
      - **Pass Protection ID** - how was this player involved in pass protection, if at all?
      - **Hand on Ground ID** - did defender begin play in a stance w/ hands on ground, or upright?
      - **Gap Rushed ID** - between which 2 players did defender attempt to reach the QB?
      - **Off Positioning ID** - offense position codes
      - **Def Positioning ID** - defense position codes

      - **Press Jam ID** *(unlabeled header, missing data?)* - for man-to-man pass coverage, what type of defense was used (press/jam)?
      - **Pass Route ID** *(unlabeled header, missing data?)* - what kind of pass route offensive player ran
      - **Pass Route Yards** *(unlabeled header, missing data?)* - how many yards did offensive player gain
    - **Man Covered ID** *(missing data?)* - presumably player ID of receiver the defender was covering
- **Single Team Season Aggregate** (x1 in sample)
  - Data on one team for an entire season.
  - Each row describes one play in one game.
  - This seems like the mother lode for
  - For each play (skipping some)
    - **Period** - quarter of game (1-4)
    - **Is TwoMin?** - inside two minute warning (close to end of half)
    - **Score Time**
    - **Yd From Goal** - how far down the field remains to endzone
    - **Event Code** - see CFB_DECODE
    - **Poss Change?** - did ball switch from one team to another?
    - **Scrimmage?** - differentiates plays from scrimmage (regular plays) with unusual plays like kickoffs I think
    - **Loose Ball?** - did anybody fumble (drop) the ball on this play?
    - **Penalty Code** - penalties assigned if any
    - **Yd Gained** - yards gained on play
    - **Passer/Rusher/Receiver/Penalized Player/Off Player N/Def Player N** - see PLAYERS$2016
    - **Yards After Catch** - after catching pass, how far did receiver run
    - **Def At Line** - # of defensive lineman at line of scrimmage
    - **Form at Snap** - see CFB_DECODE
    - **Why No Catch -** see CFB_DECODE
  - ...tons of other detailed stuff I wont bother to enter here about specifics of formations, who beat who, who did what where and for how far, who made/missed tackles/etc.

- **Lookup code tables** (x9 in sample)
  - **PLAYERS$2016, TEAMS$2016** - simple lookups for player IDs ad team IDs to name, uniform number, etc.
  - **GAMES$2016 -** lookup for game IDs to specific game including some environmental parameters. note that these are per-game not per-play, so who knows if things like temp are min/max/mean/etc. over the 3 hour game. Params include:
    - **Game Type** - Regular season game vs. Post Season game (i.e. playoffs)
    - **Home/Away Team** - who was home team and who wasn.t
    - **Surface** - field surface material, e.g. grass/artificial.
    - **indoors?** - indoor stadium or outside?
    - **Temperature** - 18-101 in sample
    - **Humidity** - empty in sample
    - **Clouds** - 0-8 in sample
    - **Wind Direction** - 1-19 in sample

- **Wind Velocity** - 0-29 in sample
- **Attendance** - crowd size
- **Time of Game** - assume PM if time is 3:15 🙂
- **DECODE** - lots of general lookups here for role/route/play type names.
- **Participation** - codes for how defensive linemen line up before the play, where their opposing offensive lineman is, their stance and role in a play
- **HandoffType** - what kind of offensive play the other team ran on the play
- **Personnel** - what offensive players were in the backfield (running backs and tight ends)?
- **QB_At_Pass** - what was QB doing when he threw the ball? moving/stationary
- **WR_Bunch** - when wide receivers line up near edge of field, how many and in what formation?

Defense Analysis Suggestions

Will need to ponder more, but broadly speaking this seems like an opportunity for:

post analysis

- analyzing matchups to see what defensive players perform well regularly against particular formations or personnel types (and not well)
- look at performance across game - does player do better early on, late in game, under pressure?
- is a player more effective in certain locations and can that be leveraged?
- practice schedule prioritization - what players need to work on especially considering what's coming in upcoming game

predictive analytics

- is the other team predictable in what kinds of defensive formations they choose under certain conditions/against certain formations?
- what is the other team likely to do in response to particular game situations? when can we be more certain of that prediction?
- if we see players X Y and Z enter the game on other team, what plays are we expecting to see?
- how does this player perform in comparable environmental conditions (i.e. does he do poorly in cold night games)?

Offense Analysis Suggestions

- grass vs turf and the type of plays that yield more yards or fewer downs to score
- outdoor stadiums - analysis of play types vs temp/humidity, wind direction/velocity - and if it changed halfway through game because of playing in the other direction ... do plays need to change based on wind direction
- success of certain types of passes based on wind speed/humidity - train QB to throw certain ones (type of parabola, distance, toss type) in certain conditions