(POSTER) Freshman Engineering Student Engagement through Sport's Data Analytics

Adam Tennant
atennant@usi.edu

Follow this and additional works at: https://docs.lib.purdue.edu/aseeil-insectionconference

Part of the Engineering Education Commons, and the Sports Studies Commons

Freshman Engineering Student Engagement through Sport's Data Analytics

Adam Tennant, Ph.D., P.E. Assistant Professor of Engineering
University of Southern Indiana Department of Engineering

Introduction

Instructors of engineering for the last quarter of a century have attempted to engage undergraduates through project based learning. Often these topics can fall short due to a lack of interest by the students. Faculty led student research experiences addressing pop-culture topics of interest can often stimulate student engagement and have the possibility to positively impact student-retention rates.

This presentation reports on the preliminary findings of a sport's data analytics approach for faculty-led student investigations and increased engagement. Freshman engineering students will rank teams or individuals through the PageRank algorithm creating a model of the complexity of sports. The specifics of the research will be briefly discussed in this presentation along with a general feel of engagement before and after the project.

This presentation includes two semesters of findings and observations from the fundamentals of engineering course (ENGR 107) that all freshman-engineering students are required to take at the University of Southern Indiana. The course focuses on engineering problem-solving methods with much of the course dedicated to formulating problems for computer solutions using MATLAB or Excel.

Students often come in with little coding experience and often become frustrated and can be verbally heard questioning if they chose the right major while working. Utilizing an interesting topic such as sport's data analytics is an attempt to engage students in the learning process and motivate them through difficulties.

Spring 2018: In NCAA (National Collegiate Athletic Association) basketball, Division-1 (D1) teams compete in regular season games to be given a chance at playing in the final tournament of the season, the NCAA Men's Division 1 Basketball Tournament (i.e., March Madness, Big Dance). Each year, thousands of people fill out a bracket in an attempt to predict the outcome of the tournament. Most individuals simply guess based on a very limited knowledge of the regular season performances of each team within the tournament. With an average of 5832 games per season between 640 teams, it is impossible for any person to watch every game every team plays. The instructor and class will explore whether Google's PageRank algorithm is sufficient for two purposes. The first purpose is to provide an unbiased alternative method for deciding which teams will participate in the tournament. The second will be to find a more accurate way to predict outcomes in the NCAA tournament.

Spring 2019: No other sport captures the imagination and is as fundamentally understood around the world as boxing. Although the science of boxing for the combatants is brutally simple, to hit and not get hit, the sport additionally offers a rich source of data for complexity science. Boxing, with its long history, has a depth to its data that can be explored with complexity science to yield insights that analysis through standard statistics has left in the dark. The instructor and class will explore whether Google's PageRank algorithm is sufficient to create an all time greatest list regardless of weight a so called pound for pound list.

Materials

Software Requirements
Microsoft Excel: Used for preprocessing data set to put in a source target arrangement to later build a directed graph. After the PageRank scores were calculated excel was used to sort the results.

MATLAB: Used for creating the directed graph from the data set and performing the PageRank calculations

Gephi: Planned to be used to create high quality images of the directed graph incorporating the PageRank scores.

Data Sets
Boxing: BoxRec is an online database that is the go to source for boxing data post Marquess of Queensberry Rules. Both male and female, past and present professional boxer's records are cataloged on this website for the public to explore. There are in excess of over 2.1 million bouts entered into the database that have been researched and validated by global volunteers. Besides the results, the boxing data bout history contained on the site can include venues, referees, judges, official weights, knockout times, promoters, and descriptions of many bouts. The data set although publically posted is not available for download. The instructor approached the staff of the website to obtain data who graciously agreed to send the data set in .csv format.

Basketball: Spreadsheet-Sports is a platform that contains a collection of data sets from a variety of sports both on the collegiate and professional levels. The site is built for the sports analytics enthusiast and offers links to download data and tools to analyze. The NCAA Division One basketball data set for each season is publicly made available prior to the start of the tournament. The data set includes results, venues, points scored, and dates; all in a heavily formatted Excel Spreadsheet that is user friendly.

Teaching Tools
VoiceThread: Used to create instructional videos so that students can have detailed directions of steps involved to perform the analytics. This allows them to work at their own speed and to review later for clarification.
Blackboard: Instructional and Informational videos are posted here along with data sets that cannot be downloaded. The project is also broke down along assignments posted on Blackboard including grading rubrics.

Methodology

Methodology for Ranking Boxers (Spring 2019)
1. The instructor introduced the students to boxing by taking a historical perspective emphasizing the cultural significance of the sport. The students then explored the BoxRec website.

2. The instructor acquainted students with the PageRank Algorithm by the story of Google's founding and impact on society.

3. The instructor taught the students the idea of a directed graph by a simple example of links on webpages. Then expanded the idea of a directed graph to boxing by a visual.
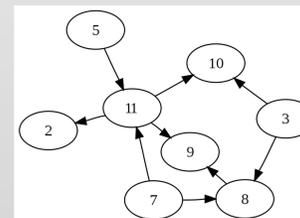
4. The instructor then gave the students the data set and explored a few entries of the bout history with them. Students gained clear understanding of how the data was not setup for the creation of a directed graph within MATLAB.

5. The instructor then began with the students to preprocess the data in Excel. Explored the idea of filtering data, cell referencing, data management, and vlookup to create a source and target arrangement .csv for the creation later of a directed graph.

6. As a class the students worked through a small example of the creation of a directed graph in MATLAB.

7. Students then were assigned separate data sets from BoxRec to create directed graphs of Boxers.

8. Explored the centrality function within MATLAB and obtained PageRank scores by using the function on the newly created directed graph.

9. Exported data set to Excel and began to process results.

10. Students prepared visualizations and began writing process for a shared journal article documenting the results.

** The same approach above was utilized prior in the spring of 2018 for basketball data set and building a March Madness Tournament Bracket.

Results
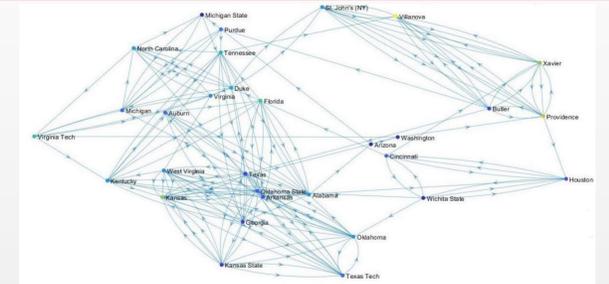
Top 64 Teams for 2018 Utilizing PageRank Analysis

Team	PageRank	Team	PageRank
Villanova	0.012255	Baylor	0.006026
Providence	0.010994	Missouri	0.005955
Xavier	0.010419	Texas Christian	0.005949
Kansas	0.010376	Louisiana State	0.005851
Florida	0.009835	Gonzaga	0.005575
Virginia Tech	0.009583	Southern California	0.00575
Tennessee	0.009303	Arizona State	0.005683
Duke	0.008784	UCLA	0.00567
North Carolina	0.00862	Saint Marys (CA)	0.005505
St. Johns (NY)	0.00853	Creighton	0.005452
Alabama	0.00837	Oregon	0.005299
Kentucky	0.008318	Temple	0.00511
West Virginia	0.008262	Stanford	0.005026
Oklahoma	0.00821	San Diego State	0.00501
Virginia	0.008038	Marquette	0.004939
Purdue	0.007887	Middle Tennessee	0.004928
Texas Tech	0.007613	Seton Hall	0.004816
Michigan	0.007591	Mississippi State	0.004764
Auburn	0.007588	Rhode Island	0.004654
Cincinnati	0.007484	North Carolina State	0.004595
Oklahoma State	0.007364	Clemson	0.004578
Houston	0.007243	Ohio State	0.004519
Arkansas	0.007166	Western Kentucky	0.004474
Texas	0.006888	North Carolina-Greensboro	0.004281
Butler	0.006788	South Carolina	0.004241
Georgia	0.006731	Loyola (IL)	0.004228
Kansas State	0.006703	Nevada	0.004183
Washington	0.006637	Louisville	0.004104
Arizona	0.006409	Colorado	0.004096
Wichita State	0.006375	Penn State	0.004002
Michigan State	0.00632	Miami (FL)	0.003995
Texas A&M	0.006273	Wofford	0.00398

Conclusion

Instructors Impression: In the spring of 2018 the students were very excited to start the project and continued with enthusiasm throughout all the way to completion. They were further challenged to add weights to the links so that all wins were not treated equally. For instance a win during a road game was worth more than a win at home and thus affected the PageRank scores. Additionally the students were assigned to code one improvement to the class project work. At the end of this semester, two students were identified to carry on the project in an independent research class. The two along with the instructor performed more of a longitudinal study of multiple seasons and have recently submitted an article to a peer reviewed journal.

During this 2019 spring semester, a much smaller number of sevens students were enrolled in the ENGR 107 class. The instructor wanted to return to a previous topic of boxing he had researched. A project has been designed with gathered data to create an all time pound for pound list of boxers. The intent again is to motivate the students to have a positive view of coding and MATLAB. This work is currently ongoing and a survey has been sent out to the students to assess if the project is working as envisioned.

References and Acknowledgements

Tennant, A. G., Ahmad, N., & Derrible, S. (2017). Complexity analysis in the sport of boxing. *Journal of Complex Networks, 5*(6), 953-963.

Matthews, N., McClain, A., Smith, C.M.L., Tennant, A. G. (submitted 1/22/2019). Application of PageRank algorithm to division-I NCAA men's basketball as bracket formation and outcome predictive utility. Journal of Quantitative Analysis in Sports.