AI Driven Computational Journalism to Generate Interesting Stories for Collegiate Sports

Jyotisman Banerjee, Drew Bertram, Yuvraj Daga, Parth Mau, Malay Rai, Yang Wang
jbanerj@purdue.edu; dbertra@purdue.edu; ydaga@purdue.edu; pmau@purdue.edu; rai36@purdue.edu; yangwang@purdue.edu

ABSTRACT

To develop a computational journalism model capable of identifying interesting trends in American collegiate athletics. The traditional method of sports journalism is time consuming, often inaccurate and ignores many significant trends. We analyzed college football data using enhanced ML algorithms to build a model that generates interesting trend stories in collegiate football in real time and helps fill the gap and limitations of manual generation of stories associated with traditional sports journalism methods.

INTRODUCTION

The fandom for football is growing exponentially. Fans consume sports digitally, have shorter attention spans and require engaging statistical trends and stories to remain hooked to the game. Sports media broadcasts insights through interactive graphics.

UNIQUE TREND STORIES → **Lead to** → HAPPY FANS | MORE REVENUE

What is a trend in sports? It is the number of times a player or team achieves a specified statistic in a game over the total number of games played.
Example: Carson Strong of Nevada had 290 or more passing yards in 9 of his last 12 games

Current State of identifying unique trends has **4 MAJOR PAIN POINTS**

- Manual, Inefficient, Tedious
- Neglects many unique trends
- Media lack access to cross-sport analytical model
- No Live Trends (Only pre/post game)

RESEARCH OBJECTIVES

We attempt to solve the current pain points by working on the following objectives :
1. To develop **AI Driven Computational Journalism Model** capable of identifying trends.
2. A model to identify what trends within the data would be **unique** to the audience American college sports media.
We reviewed literature about – Computational journalism, analytics in sports and media, anomaly detection, value of trend and streaks in sports. Post this, we imbibed the use of statistical modeling in sports journalism, computational journalism in various sectors, and anomaly detection into our model.

METHODOLOGY

The final output of our algorithm is an **interesting story sentence**.

Entity (Player/Team) — Relevant Stat — Event Total
Carson Strong has thrown **4 or more Touchdown Passes** in **3** of his last **5 games**
Qualifier of Stat — Event Success — TREND

BREAKDOWN OF METHODOLOGY

Step 1 : Identifying the Entity and Stat

ENTITY / STAT ← Input by User OR Latest Stat Auto-selected

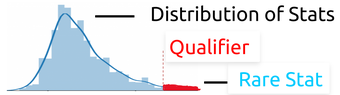
Example: *"**Malik Willis** had **200 Passing Yards** in a game"* (Entity: Malik Willis; Stat: 200 Passing Yards)

Stat picked according to player position

Step 2 : Set the Ideal Stat Qualifier

Stat -> Qualifier -> Success -> Rare Stat

(Passing Yards >200 / Passing Yards >175)



Step 3 : Identifying Rare Trends

Entity	Stat	Qualifier	Last Game #	Success
Malik Willis	Passing Yards	>200	1	1
			2	1
			3	0
			4	1
			5	1
			6	0
			7	1

Algorithm finds **Rare Trends** in the stat

Passing Yards > 200 Yards

4 /5 Games ■ 1/2 Games ■ 2/5 Games

RARE TREND! — Who cares?

Step 4 : Develop Context for the Story

Entity	Stat	Qualifier	Success	Total	Occurrence Rate	Level	Last Time - Entity	Last Time - Season
Brian Robinson Jr.	Rushing Yards	>100	7	9	16%	All Season - Conference	Derrick Henry	2015
Aidan O'Connell	Passing Touchdowns	>3	5	7	19%	All Season - Team	Joey Elliot	2008
Purdue	Field Goals	>3	6	7	9%	This Season - Conference	NA	NA

Algorithm adds significance to the unique trend. In what context is the trend unique? When did this last happen? How rare is the trend?

RESULT – Final Outcome

Using the Methodology, we got our final output – **THE INTERESTING STORY.**
1. **Purdue** has **more than 3 field goals scored** in **6** out of its last **7** games.
Last Occurrence: **NA** in **NA**.
2. Brian Robinson Jr has had **more that 100 rushing yards** in **7** out of **9 games**.
Last Occurrence : **Derrick Henry** in **2015** Season.

Broadcasters can use the algorithm to engage fans with unique stories.

CONCLUSION

With the evolution of technology in the last decade; collection of data has become more accessible, and this has facilitated the growth of sport analytics. Sports analytics is gaining popularity in mainstream sports culture. With our model, members of the sports media will receive trend updates in real time. This will allow sports media to present these trends on social media platforms, broadcasts, and print media to connect with and **grow their audience.**

FUTURE SCOPE

In its current phase, the algorithm is designed for collegiate football at its helm. Considering its dynamic nature, it can easily be extended to incorporate different sports such as soccer, baseball, basketball, amongst others.

Lastly, we believe that there is a need to personalize the final sentence structure. Research in natural language processing is required to ensure proper formatting for team/player as entities, and a special focus on different types of stats (offense, defense stats, sport specific)

ACKNOWLEDGEMENTS

We'd like to thank the teaching staff of Krannert School of Management, Purdue University; especially Dr. Wang and Dr. Lanham for their continuous support during the project. We are also very grateful to our client partner Dr. Shastry for giving us the opportunity to work on this project and his unparalleled guidance throughout.