The Analytics Edge, Final Report:
Building an Article Recommender for Bleacher Report Basketball

for Professor Dimitris Bertsimas and Emma Gibson

by Stephen Albro & Cyrille Combettes
salbro@mit.edu, cyrille@mit.edu

May 7, 2018

1 Abstract

2 Motivation and Project Abstract

As avid basketball fans, we have enjoyed reading a variety of NBA articles to keep up-to-date on the league during our busy year in the MBAn program. Bleacher Report, a primary digital destination for sports article readers, contains well written articles covering the league, and competes as advertisement real-estate with the likes of ESPN.com, FoxSports.com, CBSSports.com, SBNation.com, and others.

In an age of cut-throat digital competition, it is vital for websites to be able to retain visitors. The hope for a Bleacher Report article reader is two-fold. First, that the visitor jumps from article to article *on the website*, rather than returning to the search engine results page. Second, that the visitor falls in love with the content and coverage of Bleacher Report's articles, and develops a loyalty to the website.

We feel that having a micro targetted article recommendation system would provide Bleacher Report with an analytics-based edge over its competitors, since well targeted articles could increase the chances of prolonged visits, and increase website loyalty by making users feel known. In this paper, we develop a prototype for an article recommendation system using state of the art natural language processing techniques for both cleaning and analyzing bleacher report articles. Specifically, we aim to cluster over 3000 Bleacher Report articles by testing three different approaches: Topic Modeling, Word2Vec text embeddings, and NMF (CYRILLE WHAT DOES THIS STAND FOR). {GO INTO MORE DETAIL HERE ABOUT HOW CLUSTER HELPS THE RECOMMENDER. }

3 Data Collection and Cleaning

- scraping
- NLP cleaning (removing stop words, etc.)

4 Methods

- include tuning and cross validation, etc.
- include tables, cool visualizations

4.1 ¡METHOD 1¿

4.2 ¡METHOD 2¿

4.3 ¡METHOD N¿

5 Discussion

5.1 Which Approach is Best

- how to we measure success, etc.

5.2 Why This Project Was Important

- we learned a lot about all three methods - most of the world's data is text
- lots of online advertising
- lot of research in NLP because anlytics provides an edge