

---

# A Come-from-Behind Win or a Blown-Save Loss: Perspectives in Baseball

---

Alice Oh

AOH@CSAIL.MIT.EDU

MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge MA, 02139 USA

## 1. Introduction

One can say a baseball game was a come-from-behind win or a blown-save loss. It all depends on who you are and how you view the event. The goal of our HiT (Highlights in Two) Project is to recreate a multimedia highlight summary of a baseball game from different perspectives, such that the users can view the event with their preferred bias. As the first step in automatically generating biased multimedia highlights, this paper presents the results of the analyses of bias in written media.

As an example of bias in the media, let us look at how media describes the fifth game of the American League Championship Series (ALCS) of 2004 between the Boston Red Sox and the New York Yankees. The Boston Globe and the New York Times featured headlines, respectively, “Two wins in hand, two to go”, and “Yankees lead series, 3-2”. The unbiased fact was that the New York Yankees had won the first three games of the best-of-seven series, then the Red Sox won the next two games. The biased interpretations were that, for a New Yorker, the Yankees were still up by one game, but for a Bostonian, the Red Sox had just won two games. The divergence of perspective does not end there. What follow the headlines are two articles that hardly seem to describe the same event. Besides these two game wrap-up articles, the two newspapers, as well as other online sports websites, such as ESPN ([www.espn.com](http://www.espn.com)) and MLB ([www.mlb.com](http://www.mlb.com)) feature many articles presenting in-depth analyses from a variety of perspectives.

As such, sports fans have many choices for written media, but they have very few choices for visual media. Local and national sports channels broadcast actual games and/or highlights, but they only offer one perspective. The Internet offers more flexibility, such as MLB.com where users may search for short video clips, but those discontinuous and out-of-context clips do not satisfy the user’s needs for a biased analysis or an insightful story of the game. The HiT Project is directly aimed at those needs of the users.

## 2. Related Work

There have been many recent successes in multimedia analysis and modeling. In the sports video domain, research

systems successfully analyze videos of soccer (Xie et al., 2004), baseball (Zhang & Chang, 2002), and other sports. The outputs from these systems vary in the level of domain specificity and semantic detail, but a typical classification algorithm produces an index of all events in a sports video such as play/break, score/batter change/base change. Such successes naturally lead to the next step of developing a system for automatic highlight generation.

Tailoring multimedia summaries to a specific user based on a user preference model has been an active area of research, particularly in the broadcast news domain (cf. (Maybury et al., 2004)) where they retrieve news stories that are of interest to a user. Our problem is different from the broadcast news domain in that the individual events we retrieve for the user need to be connected together to make one continuous story. Babaguchi et al. (Babaguchi et al., 2004) have done extensive research on generating video abstracts based on personal preferences, but our system goes beyond theirs in two ways. First, we infer the significance of the events based on context-dependent semantic features. For example, in a high-scoring baseball game, a one-run home run in the first inning may not be significant, but a two-run walk-off single in the ninth inning is probably much more significant. Secondly, we focus on the overall story that matches a user’s perspective, as opposed to selecting individual events based on user preferences.

## 3. Biases in Media

Although sports journalists stay true to the facts when they report on baseball games, many local newspapers and television broadcasts spin the stories to cater to the local audience. To understand how they actually do this, we analyzed articles from several local newspapers and compared them among themselves and against the Associated Press (AP) articles to quantify how “bias” is realized. We looked at three articles per game for six games: April 20, 25, 26 between the Boston Red Sox and the Baltimore Orioles, and May 13, 14, 15 between the Red Sox and the Seattle Mariners. We used articles in the Boston Globe for Red Sox coverage, Baltimore Sun for the Orioles coverage, and the Seattle Times for the Mariners coverage. We used the AP articles as the neutral counterpart.

Newspaper (Game)	Bos	Bal	Bos (Norm)	Bal (Norm)
Sun (3 Games)	91	128	0.83	1.52
Globe (3 Games)	145	52	1.33	0.62
AP (3 Games)	109	84	1.00	1.00

Table 1. Word counts of player and coaching staff names.

### 3.1 Bias in Discussing Players

The articles showed significant differences on three dimensions. First, the local articles covered more content on their home team’s players compared to the AP articles. This is evidenced by the difference between the word count of the home team player names and the word count of the other team player names. Table 1 shows the comparison of the word counts between the Baltimore Sun articles and the Boston Globe articles. The first two columns show the raw counts, and the next two columns show the counts normalized to the counts in the AP articles. To save space, we present only the total counts, but all three games display consistent differences. The counts and the normalized counts clearly show that the Baltimore Sun writes more about the Orioles players, and the Boston Globe writes more about the Red Sox players. The Seattle Times articles and the Boston Globe articles between the Mariners and the Red Sox showed similar biases. A chi-square analysis shows significant difference between the Globe and the Sun at  $p < 0.001$ , also significant differences between the AP and the Globe, AP and the Sun, both at  $p < 0.01$ .

### 3.2 Bias in Choosing Events

Second, the local papers discussed more of their home team’s successes than the neutral AP articles. To see this, we can dissect the events of each game into three categories: events that are included in all three articles, events in the home team’s articles, and events in the visiting team’s articles. For a given event, we can say it is a *local-positive* if it is an offensive play that led to, or could have led to scoring (e.g., a hit, a walk), or if it is a defensive play that prevented the other team from scoring, (e.g., a strike out, a double play). On the ten games we analyzed with three articles each, the home team’s newspaper discussed 17 local-positive events on average, compared to 12 local-negative events. The AP articles did not show any significant difference in their coverage of one team versus the other. A t-test shows a significant difference between the Globe and the Sun articles at  $p < 0.01$ .

### 3.3 Bias in Interpreting Plays

Third, the local articles describe the events of the game with more focus on the home team’s players. This bias is difficult to quantify because it is often reflected in the subtle nuances of the language used. The following shows one of

the more straightforward examples of how a play was described by the three articles. The first is from the AP article and shows the objective account of two Seattle home runs. “Richie Sexson and Raul Ibanez added solo home runs for Seattle in the third inning, the first time the Mariners hit consecutive homers this season.”

The Boston Globe, interprets the same home runs as the pitcher’s mistakes. “Gonzalez gave up back-to-back home runs leading off the third to Sexson and Raul Ibanez.”

The Seattle Times attributes the home runs to the batters’ performance, further elaborating with the appropriate statistics and highlighting the Mariners’ lead. “In the bottom of the inning, Sexson led off with his 10th homer, putting him on pace for 46 for the season, and Ibanez hit his fifth for a 5-4 Seattle lead.”

Although there are many ways to express the different interpretations, the most frequent ways are putting the home team’s player as the subject, providing statistics that highlight the home team player’s strengths, and mentioning scores when home team has captured the lead or has caught up to within a small margin.

## 4. Future Work

The next step is to apply the analyses in creating a system for generating highlights. As an intermediate step, we have also collected audio highlights from local radio stations and plan to do similar analyses. We have built a prototype system, and a demo will be shown at the CSW.

## References

- Babaguchi, N., Kawai, Y., & Kitahashi, T. (2004). Personalized abstraction of broadcasted american football video by highlight selection. *IEEE Transactions on Multimedia*, 6, 575–586.
- Maybury, M., Greiff, W., Boykin, S., Ponte, J., McHenry, C., & Ferro, L. (2004). Personalcasting: Tailored broadcast news. *User Modeling and User-Adapted Interaction*, 14, 119–144.
- Xie, L., Xu, P., & Chang, S. (2004). Structure analysis of soccer video with domain knowledge and hidden markov models. *Pattern Recognition Letters*, 25, 767–775.
- Zhang, D., & Chang, S. (2002). Event detection in baseball video using superimposed caption recognition. *Proceedings of ACM Multimedia’02* (pp. 315–318).
- Zhou, W., Vellaikal, A., & Kuo, C. (2000). Rule-based video classification system for basketball video indexing. *Proceedings of ACM Multimedia Workshop* (pp. 213–216).