Using Twitter to Detect and Tag Important Events in Live Sports

James Lanagan and **Alan F. Smeaton**
CLARITY: Centre For Sensor Web Technologies
School of Computing
Dublin City University
Dublin, Ireland

Abstract

In this paper we examine the effectiveness of using a filtered stream of tweets from Twitter to automatically identify events of interest within the video of live sports transmissions. We show that using just the volume of tweets generated at any moment of a game actually provides a very accurate means of event detection, as well as an automatic method for tagging events with representative words from the tweet stream. We compare this method with an alternative approach that uses complex audio-visual content analysis of the video, showing that it provides near-equivalent accuracy for major event detection at a fraction of the computational cost. Using community tweets and discussion also provides a sense of what the audience themselves found to be the talking points of a video.

1 Introduction

As the web moves further and further away from traditional publishing models, more information is being generated by different sources every day. Alongside professionally produced news broadcasts and interviews we are now able to follow real-time streams of information, the classic example being Twitter (*http://www.twitter.com*).

In this paper we focus on the exploitation of Twitter information in the sports domain. Sports is particularly suited to this sort of analysis since there is an ever-present and active audience for all sports. By analysing the content of their conversation (both volume and vocabulary) we show that it is possible to achieve very good event detection and classification within sports video through faster but equally accurate methods than audio-visual analysis alone. We also show that we can display what the audience *themselves* found to be the most interesting and exciting moments; this is not possible using traditional audio-visual approaches.

This paper is organised as follows: in Section 2 we discuss work from both the content analysis and user-generated content analysis fields, each of which is important in grounding our own work. Section 3 describes the motivations behind our research, highlighting the benefits of introducing user-generated content (Twitter) into our own framework and processes. The methodologies and implementation we developed are outlined in Section 4 before presenting the results and outcomes of our experiments in Section 5. Finally, we summarise some of the problems we have encountered during this work and our aims for extending this research in Section 6.

2 Related Work

Past research has shown that using the audio and visual features of video to discover events of interest within video of different field sports (Lanagan and Smeaton 2007; 2009; Sadlier and O'Connor 2005) is highly effective. Complementary to this is the approach of (Ferguson et al. 2009) who attempt to introduce an element of sociality to the sometimes solitary pastime of television viewing.

Folkonomies arose as a possible solution to the problem of describing content within webpages There have been attempts to transfer lessons learned within folksonomy research to Twitter (Wagner and Strohmaier 2010).

Twitter allows its users to publish updates or 'tweets' about any topic they like, but limits the length of these tweets to 140 characters. The Twitter API allows for the focussed search of this public stream based on issued queries. Similar to standard information retrieval systems like Google, Bing, and Yahoo!, all tweets returned in answer to these queries contain the query terms or hashtags [1]. We filter the Tweets from the public timeline by hashtags and only retrieving those relevant to each match.

Twitter has already been used to show large correlations between its tweeting population and the population and attitudes of the real world (Sakaki, Okazaki, and Matsuo 2010). Shamma *et al.*'s work (Shamma, Kennedy, and Churchill 2009) is closest to our own: they use the volume of tweets from people watching the first presidential debate in 2008 as an indication of the level of interest. Their research centred around a live televised event much the same as our research. The event itself however has an important difference; political debate can give rise to heated conversation and discussion. With sports we have found that interest (as signalled by the volume of tweets) is more intermittent, centred on the key moments from within the sports event. We consider any event that changes the score of the sports event, or bookings or other disciplinary actions by the referee(s) as important[2].

[1] Hashtags in Twitter (#keyword) are used to make searching and collecting tweets about a common topic easier.

[2] We use the match reports given by ESPN Sports Centre:

3 Motivation

The detection and easy navigation of highlights within sports media is something that is both appreciated and desired by viewers (Lanagan and Smeaton 2007). The combination of both audio and visual video content features has been able to successfully detect many of the most important events within a match (Sadlier and O'Connor 2005)[3].

Our main motivations for using Twitter as a source of information are speed and enrichment. By looking at the words that people are using within their tweets, we are able to create tags that best describe and classify the highlights to which they are temporally aligned. Performing audio-visual analysis is time-consuming; Twitter provides a real-world approximation of the interest, allowing for the real-time segmentation and display of highlights.Using the Twitter stream alone however ignores any of the content features themselves, leading to disorientating initial frames that appear to jump into the middle of the action. To counteract this, we combine the tweet information with a video shot-boundary detection algorithm (SBD) so as to provide more intelligently bounded highlight videos.

4 Experiments

We asked 6 annotators to each mark up the events of interest within 3 of the videos we recorded (all of these videos are of soccer matches, though we speculate event descriptions for rugby, American football, or other field sports may prove slightly easier due to the stop-start nature of those sports).

Table 1: Significant event boundaries and durations in seconds as marked up by our annotators.

	Boundary	*Std. Dev*	*Min/Max*	*Duration (Secs)*
1	start	3.93	1288	
	end	0.84	1350	62
2	start	4.02	5584	
	end	8.99	5688	104
3	start	3.27	3995	
	end	8.77	4080	85
4	start	3.78	2730	
	end	1.60	2780	50
5	start	2.48	5617	
	end	7.71	5730	113
6	start	3.78	6347	
	end	3.39	6414	67
7	start	4.36	945	
	end	6.50	1028	83
8	start	48.6	2390	
	end	30.2	2550	160

Table 1 shows the event boundary decisions in seconds made by our annotators for the 8 goals scored during the 3 matches. The durations shown are the maximum event spans using the least/greatest values for start/end second. We feel it is better to show too much of the original video during a highlight than lose some of the highlight content itself.

We use the method of hashtag filtering – as explained in Section 3 – to construct and retrieve filtered tweets about the football matches we recorded, 4x soccer and 4x international rugby. The numbers and frequency of tweets per game vary greatly as shown in Table 2, with one of the soccer games having far less than any other game. This was a far lower profile game than all the others, yielding some interesting results and problems in terms of event detection.

Table 2: Tweet statistics per game for the 4 soccer, and 4 rugby matches.

Game	*Tweets*	
	Avg/Min.	*Total*
ACM-ManU	19.76	1877
Cov-Por	1.03	101
Liv-Rea	32.3	4073
ManU-ManC	12.95	1256
Eng-Ire	33.42	3108
Eng-Wal	35.30	3142
Fra-Ire	27.53	2615
Wal-Sco	23.86	2291

While there is no set convention for designating hashtags to tweets within Twitter in general, one has started to grow around sports-related tweets. This was highlighted during the World Cup 2010 when Twitter had their own hashtag-filtered pages for each team, game, and overall competition.

We have constructed complex queries using this same convention (e.g. *#ire #eng #6nations* for the 6 nations Ireland vs. England rugby football match) so as to retrieve the maximum number of relevant tweets for each sports video. In the later experiments to find suitable tags, non-English tweets are ignored. In this research we are not interested in any tweets that occur outside the bounds of actual live gameplay including tweets during half-time or extra-time intervals.

The aim of the process here is to be able to provide an end-user with full highlights of a match through their mobile device, or as part of a larger system (Lanagan and Smeaton 2009). As with previous systems, we look to combine simple content analysis with more complex information. In the past this was complex audio-visual information (Sadlier and O'Connor 2005), providing per-second confidence values for event occurrence within the video. In our approach, we substitute this information with per-second tweet counts allowing us to use the same overall technique, dramatically reducing computational complexity.

Sports video contains many hard cuts leading to shots that vary greatly in duration. These shots alone would not provide a good basis for highlight retrieval and browsing, and so we combine the initial shot boundary detection (SBD) results with some heuristic measures. As we have seen, significant events are in the region of 60 seconds in duration. We amalgamate the shots into segments using the following

http://soccernet.espn.go.com/

[3]The detection approach used is multi-modal and relies on both audio and visual information streams to determine confidence levels. Six Support Vector Machine (SVM) classifiers are used that detect the presence of player close-ups, crowd shots, scoreboard changes, increased audio activity, playing field boundaries and increased visual activity. A more detailed description of training and testing maybe be found in the original paper.

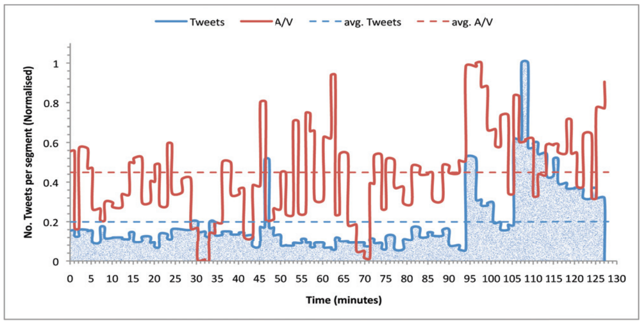

Figure 1: Normalised per-segment event confidences for each of the matches (Twitter graph shown with shading.).

procedure:

(i) Find the top 10 per-second event confidence maxima as given by either Twitter or audio-visual features.

(ii) Create 60 second windows around these local maxima.

(iii) Combine shots together until a segment is formed that is of at least 60 seconds in duration.

(iv) If an event is present, continue the segment until the event has completed.

(v) If two events are within 30 seconds of each other, combine the events and do not end the segment until both events are contained.

(vi) Find the event confidence for the segment by calculating the mean event confidence of every second contained within the segment.

(vii) If the video is not complete, return to Step (iii).

We choose to return only the 10 most significant events within a video, or as many events that are found with higher confidence values than a given threshold.

The dynamic threshold is now chosen in a second pass. We average across entire segments rather than across minutes etc., since minutes are no longer the measure of interest.

Figure 1 shows the normalised per-segment event confidences for a recorded matches matches were recorded between late January and early March 2010, and show a wide range of sporting competitions.

4.1 Twitter as Reaction

It is important to consider the difference between audio-visual and Twitter information as interpretations of event occurrences. The 6 SVMs used to provide event confidence values do so by looking for increases in audio and visual excitement. While this excitement is a result of crowd and editors' reactions to events within the game, the reaction is almost instantaneous; this is not true of Twitter. We can not look to the segments with the highest confidence values in the Twitter combination for the presence of an event, but instead for the near-immediate reaction to an event. We therefore consider events to begin at the start of the segment preceding a local maximum, and end when that local maximum's segment ends. While audio-visual features such as field end-lines and increased crowd noise may be used as indicators of an event about to happen, there are no indicators

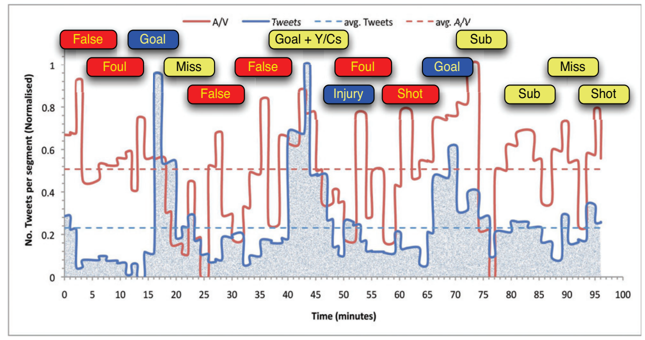

Figure 2: Comparison of events captured by each combination. Yellow events found by both algorithms, red and blue found by their respective algorithms.

in the Twitter data until after the event is finished – in this sense Twitter is truly reactionary.

Table 3: Top terms occurring for given events within different games. Occurrences are in brackets

FRA-IRE:Penalty	*ManU-ManC:Goal*	*WAL-SCO:Try*
#6nations (44)	#manutd (50)	#6nations (299)
flannery (19)	**penalty (46)**	wales (90)
#rugby (17)	1 (44)	scotland (62)
lucky (8)	#mufc (42)	#rugby (56)
ireland (6)	#mancity (34)	#wales (44)
s (6)	**tevez (30)**	24 (44)
penalty (5)	#mcfc (29)	game (38)

When determining tags for the events we remove all non-English tweets as these are considered noise for tag-generation purposes. We then group all tweets that occur within the event boundaries, and generate a 'bag-of-words' from these tweets. We ignore any term that occurs less than 3 times in an events bag-of-words. Highly-frequent terms are removed using the WordNet stoplist[4]. We can see from Table 3 that there are still many words left within the corpus that are in no way discriminatory (i.e. is, was, and single letters). Removal of all words that are not of some minimum length, and Hashtags should perhaps have been performed as they have no discriminative value.

5 Results

The combination of Twitter filtered search result with SBD content analysis appears to be very effective in detecting significant events. Not only this but it also provides more accurate measures of the actual significance of the events. Table 4 shows the comparison of the two techniques' abilities to find all ground-truthed events within our three annotated soccer matches.

We can see that the events missed by the Twitter algorithm are bookings, but all goals are found. The segmentation technique also succeeds in finding all of an event (length

[4]http://www.d.umn.edu/tpederse/Group01/WordNet/wordnet-stoplist.html

Table 4: Ground truth events found consisting of goals(G) and yellow cards(Y/c). Events missed are in brackets.

Game	*Twitter*	*A/V Features*
Cov-Por	2 G (2 Y/c)	All (Many False Pos.)
Liv-Rea	3 G, 1 Y/c (3 Y/c)	All (Many False Pos.)
ManU-ManC	3 G, 2 Y/c (1 Y/c)	2 G, 3 Y/c *(1 G)*

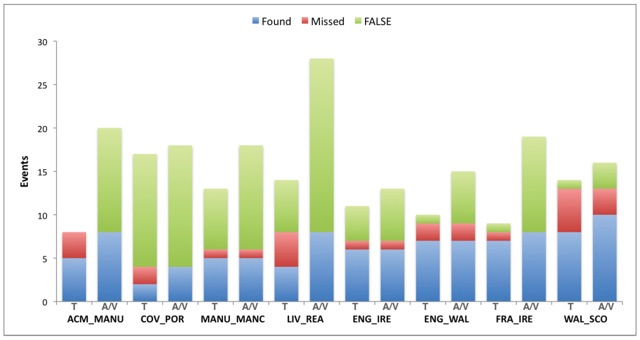

Figure 3: A comparison of event detection accuracy across the 8 games, and 2 techniques.

in seconds) as annotated by our annotators, and not truncating it due to simple tweet volume. The audio-visual feature combination is slightly more effective in finding events within the video, but this effectiveness is tempered by the large number of false positives found. It should be noted that because of the false positives that are found, not all of the significant events will be in the top 10 event listing.

Looking at the effectiveness of the Twitter-SBD combination to detect events across the remaining matches, we find that it has a 100% accuracy rate in finding goals in soccer matches overall, finding all 5 goals in the ACM-ManU match. For rugby it only misses 2 of the 18 tries in the 4 rugby matches, both of which occur in quick succession to another significant event. The detection rate for penalties within the rugby matches is slightly lower at ∼80%, these occurring more frequently. With regards false positives, the audio-visual algorithm finds far more in the rugby matches than in the soccer matches.

6 Future Work

One of the major issues we encountered was a reliance upon conversation and tweet volume as a principle input; without tweets the analysis is not possible. it is also true that the number of tweets over the course of a game can significantly affect the performance of our algorithm. Another problem is highlighted in Figure 1. The level of excitement and conversation in general rose after a late goal, introducing an ‘event shadow’ as can be seen in the last quarter of the game, and inflating the overall average throughout the game.

To ameliorate both of these issues we could introduce some form of smoothing or windowing. A second approach is to combine the outputs of both of our approaches to give a combined confidence value for each event. This would have the advantage of being less reliant on the number of tweets, but the added computational complexity of the audio-visual approach was specifically what we are attempting to avoid.

7 Conclusions

In this paper we have shown that it is possible to use the publicly available tweets about a sports event to aid in event detection and summary generation for associated media. Using a filtered stream of tweets we can identify and tag the most significant events within 2 different field sports with a high degree of accuracy and success. We believe that the power of our approach lies within its seeming simplicity, allowing for equivalent results to far more computationally complex, and time consuming approaches.

Acknowledgments

This work is supported by Science Foundation Ireland under grant number 07/CE/I1147.

References

Ferguson, P.; Gurrin, C.; Lee, H.; Sav, S.; Smeaton, A. F.; O’Connor, N. E.; Choi, Y.-H.; and Park, H. 2009. Enhancing the functionality of interactive tv with content-based multimedia analysis. In *CBTV 2009: Workshop on Content-Based Audio/Video Analysis for Novel TV Services*, 495–500. San Diego, California, USA.: IEEE Computer Society.

Lanagan, J., and Smeaton, A. 2007. SportsAnno: What Do You Think? In *RIAO’2007: Proceedings of the 8th Conference on Information Retrieval and its Applications*.

Lanagan, J., and Smeaton, A. F. 2009. Query Independent Measures of Annotation and Annotator Impact. In *ESAIR ’09: Proceedings of the WSDM ’09 Workshop on Exploiting Semantic Annotations in Information Retrieval*, 35–38. Barcelona, Spain: ACM.

Sadlier, D., and O’Connor, N. 2005. Event Detection in Field Sports Video Using Audio-Visual Features and a Support Vector Machine. *IEEE Transaction on Circuits and Systems for Video Technology* 15(10):1225.

Sakaki, T.; Okazaki, M.; and Matsuo, Y. 2010. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In *WWW ’10: Proceeding of the 19th International Conference on World Wide Web*.

Shamma, D. A.; Kennedy, L.; and Churchill, E. F. 2009. Tweet the Debates: Understanding Community Annotation of Uncollected Sources. In *WSM ’09: Proceedings of the first SIGMM workshop on Social media*, 3–10. Beijing, China: ACM.

Wagner, C., and Strohmaier, M. 2010. The Wisdom in Tweetonomies: Acquiring Latent Conceptual Structures from Social Awareness Streams. In *Proc. of the Semantic Search 2010 Workshop (SemSearch2010)*.