Spot the Ball: Detecting Sports Events on Twitter

David Corney, Carlos Martin, and Ayse Göker

IDEAS Research Institute, School of Computing & Digital Media,
Robert Gordon University, Aberdeen AB10 7QB
{d.p.a.corney,c.j.martin-dancausa,a.s.goker}@rgu.ac.uk

Abstract. Messages from social media are increasingly being mined to extract useful information and to detect trends. These can relate to matters as serious as earthquakes and wars or as trivial as haircuts and cats. Football remains one of the world's most popular sports and events within big matches are heavily discussed on Twitter. It therefore provides an excellent case study for event detection.

Here we analyse tweets about the FA Cup final, the climax of the English football season, for 2012 and 2013. We evaluate an automated topic detection system using a ground truth derived from mainstream media. We also show that messages can be associated with different teams' fans, and that they discuss the same events from very different perspectives.

Keywords: topic detection, Twitter, football.

1 Introduction

Twitter has been used to detect and predict events as diverse as earthquakes [1], stock market prices [2] and elections [3] with varying degrees of success. Journalists are increasingly using social media to find breaking news stories, sources and user-generated content [4]. Here, we consider the problem of detecting events from social media within the domain of football. Association Football (or 'soccer') remains the world's most popular sport. The last FIFA World Cup final (2010, South Africa) had a global TV audience of over 500 million, with more than 2 billion people watching at least 30 minutes of football during the tournament [5]. The next World Cup is scheduled for Brazil (2014), which itself is now the country with the second largest number of Twitter users [6].

Twitter is actively increasing its ties to television [7]. Major sporting events, like many live TV shows, occur at pre-specified times; they attract large audiences; and they are fast-paced. These features allow and encourage audience participation such as sharing comments and discussing the event with a focus that moves with the event itself. Facebook and other social media are also competing for access to this valuable "second screen" [8]. Thus there is a substantial and growing demand for linking social media discussions to televised events, including sports, for example to improve targeted advertising.

M. de Rijke et al. (Eds.): ECIR 2014, LNCS 8416, pp. 449–454, 2014.

2 Methods

Our system identifies phrases (word n-grams) that show a sudden increase in frequency (a "burst") in a stream of messages and then finds co-occurring n-grams to identify topics. Such bursts are typically responses to real-world events. Here, we filtered tweets from Twitter's streaming API using the teams' and players' names as keywords.

We want to find n-grams that appear at one point in time more than previously. We measure burstiness using the "temporal document frequency-inverse document frequency", or *df-idf*$_t$ (Eq. 1), a temporal variant of the widely-used *tf-idf*. For each n-gram, we compare its frequency in one slot with that of the preceding slot, where each slot contains x tweets (here, $x = 1500$ based on experiments to optimise recall). We define the document frequency df_{ti} as the number of tweets in slot i that contain the n-gram t. We repeat this for commonly-occurring sequences of two and three words as well as single-word terms.

$$df\!-\!idf_{ti} = (df_{ti} + 1)/(\log\left(df_{t(i-1)} + 1\right) + 1). \tag{1}$$

This produces a list of terms (i.e. words or 2- or 3-grams) which can be ranked by their *df-idf*$_t$ scores. We then group together terms that tend to appear in the same tweets with a standard hierarchical clustering algorithm. We define the similarity between two terms/clusters as the fraction of messages in the same slot that contain both of them, so it is likely that the term clusters whose similarities are high represent the same topic. Therefore, we merge clusters until no two share more than half their tweets at which point we assume that each one represents a distinct topic.

Finally, the clusters are ranked by the maximum df-idf$_t$ score of their terms, so that each cluster is scored according to the most representative term in the cluster. Each of these clusters defines a topic as a list of terms, which we can also illustrate with representative tweets. The whole process is fast enough to be carried out in real-time on a single PC. We recently published more details about this algorithm [9] where we compared it with alternative algorithms and benchmarks such as LDA. We also described evaluation for recall of political news events as well as sporting events.

We also attempt to identify the team that each Twitter user supports (if any). For each user, we count the total number of times they mention each team across all their tweets. Manual inspection suggests that fans tend to use their team's standard abbreviation (e.g. CFC or MCFC) greatly more often than any other teams', irrespective of sentiment. We therefore define a fan's degree of support for one team as how many more times that team's abbreviation is mentioned by the user compared to their second-most mentioned team. Here, we include as "fans" any user with a degree of two or more and treat everyone else as neutral.

3 Related Work

We now briefly review three recent studies of Twitter-based sporting event detection. These all detect sudden increases in the volume of tweets as an indication that an event has occurred; in contrast, our system detects increases in the frequency of words or phrases, which is largely independent of the total number of messages received. Thus even if the total volume of tweets is constant, our system can detect shifts in the topic of conversations in reaction to events.

A recent study also used Twitter detect events during football matches [10]. While we have similar aims, our methods are somewhat different. After collecting tweets using suitable hashtags, their system detects events by finding spikes in the volume of tweets collected. For each spike, they analyse the words of the constituent tweets and use machine learning tools to classify the event. They consider just five classes of event (goals, own-goals, red cards, yellow cards and substitutions) and compare these classifications to the official match data to evaluate their system. Finally, they classify individual tweeters as fans of one team or another by counting the number of mentions of each team over several matches, similar to our approach.

Twitter has also been used to identify events during American football (NFL) matches [11]. In that work, tweets were collected by filtering by team names and by NFL terminology. Events were also detected by finding spikes in the overall volume of collected tweets and each event was assigned to one of a fixed number of classes using lexicographic analysis. They used an adaptive window to find events of varying impact. Their system was very effective at detecting the most significant scoring events (i.e. touchdowns), but was less effective at finding"smaller" events such as interceptions and field goals.

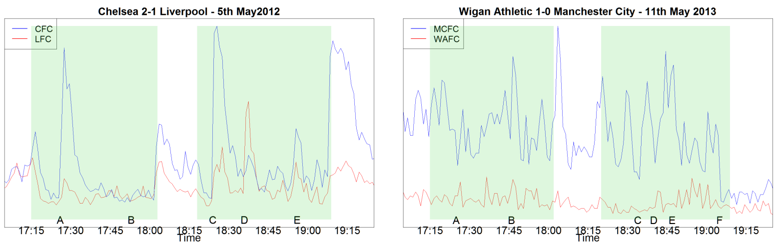
Past work on sporting event detection has often focussed on video (or audio) analysis to discover key moments. One recent study used Twitter to enhance such video annotation of both football and rugby matches [12]. They also detected events using spikes in tweet volumes in a stream filtered using hashtags. They then classified events into a few pre-specified classes and used that to create a video of match highlights.

In contrast to the above studies, our system can find arbitrary events and is not limited to pre-specified classes of event. Nor is it limited to finding just one event at a time, as we do not attempt to classify each "event minute" into a single category (unlike [10]). Our system has the potential to identify a much wider range of events than these systems, even if the events are overlapping.

4 Results and Discussion

Figure 1a shows the relative frequency of tweets from fans during the 2012 final. Both groups are active throughout the match with a number of clear spikes in activity. Chelsea fans are particularly active immediately after their team scores (points A and C) and also at the end of the match in celebration of their victory, as would be expected. Liverpool fans are more active when their team score (D). Both sets of fans are active when Liverpool nearly equalize at the end (E).

Figure 1b shows the frequency of tweets from the 2013 final. One clear feature is the large number of Manchester City tweets compared to Wigan Athletic. At the end of the Premier League season, Manchester City finished 2^{nd} while Wigan



(a) (b)

Fig. 1. Tweets from supporters of each team, FA Cup finals of 2012 (left) and 2013 (right). See Table 1 for events key.

Table 1. Key to events of matches shown in Figure 1a (left) and Figure 1b (right)

Key	Team	2012 Final Event	Key	Team	2013 Final Event
A	CFC	Ramires scores	A	WAFC	MacManaman shot, misses
B	CFC	Mikel yellow card for a foul	B	MCFC	Tevez shot, saved
C	CFC	Drogba scores	C	MCFC	Zableta booked
D	LFC	Carroll scores	D	WAFC	MacManaman chance
E	LFC	Carroll shoots, saved	E	MCFC	Rodwell free-kick is saved
			F	WAFC	Watson scores

finished 18^{th} and were relegated. Furthermore, Wigan had an average home attendance of 19,359 compared to City's 46,974 (`http://www.soccerstats.com`). This discrepancy in recent success is apparently reflected in the number of tweets. These patterns are correlated with the number of followers of the clubs official Twitter accounts. As of 28 October 2013, @LaticsOfficial has 118,512 followers; @MCFC has 1,264,369; @ChelseaFC has 2,943,118 and @LFC 2,072,077.

To evaluate our topic detection algorithm we generated a ground truth using mainstream media descriptions of the matches, primarily the BBC online commentaries. Our ground truth includes not just goals scored but also near-misses, good saves and other events likely to be of interest to fans. We define each event with a list of terms (typically 3-5) and define an event as correctly extracted only if all terms are found. We selected 13 events for the 2012 FA Cup Final (also discussed in [9]) and 15 events for the 2013 Final for this evaluation. Some of these events are listed in Table 1.

Table 2 shows the proportion of correctly recalled topics for each match as we vary the number of topics sought in each slot. As expected, as we increase the number of topics being considered, the total recall increases. For the 2012 match, recall is already nearly perfect even with just one topic, so does not increase further. For the 2013 match, a high recall is achieved when at least 4 topics are identified. This difference is due to the specific topics in the ground truth of the two matches and how many tweets discussed each topic. It also reflects the differing nature of the matches and the corresponding reports.

Table 2. Topic recall for FA Cup 2012 and 2013 sets

	Top results				
Dataset	1	2	3	4	5
FA Cup 2012	0.923	0.923	0.923	0.923	0.923
FA Cup 2013	0.333	0.600	0.667	0.733	0.733

Table 3. Examples of detected topics and example tweets for teams in both the 2012 and 2013 FA Cup finals

Dataset	**Topic detected**	**Sample tweet**
Mainstream Media Story (2012): Didier Drogba scores second goal		
FA Cup 2012 Chelsea fans	#cfcwembley #facupfinal sl chelsea goal @chelseafc	RT @chelseafc: Chelsea goal #CFCWembley #FACupFinal (SL)
	king wembley	KING OF WEMBLEY DROGBA #ENOUGHSAID #CFC
	yes drogba	Yes Drogba! #CFC
FA Cup 2012 Liverpool fans	game over #lfc	Game over #LFC sloppy playing, bad passing and conceding easy goals
	fuck #lfc	Fuck Fuck Fuckity Fuck #lfc
	didier drogba 2-0 tackling chelsea	Didier Drogba. 2-0 to Chelsea. Where was the defending? The midfield? The tackling? #LFC #FACupFinal
Mainstream Media Story (2012): Carroll (Liverpool) shoots but Cech just saves		
FA Cup 2012 Chelsea fans	petr cech	petr cech is my hero @chelseafc #CFCWembley #FACupFinal
	goal line	Hahaha! Another case for goal-line technology. #Chelsea 2-1 #Liverpool #FACup
	save #cech	what a save!!#cech#cfc
FA Cup 2012 Liverpool fans	over line ball	Ball is over the line! #ynwa #LFC #liverpoolfc #facupfinal
	goal blind	The ref n linesman is blind!!! That was a goal!!!! #LFC
	robbed #facupfinal	Robbed. Fucking robbed. #FACupFinal
Mainstream Media Story (2013): Wigan goal (Watson)		
FA Cup 2013 MCFC fans	wigan #facupfinal #wigan fucking	Fucking get in Wigan!! #facupfinal #Wigan #mcfc
	#mcfc #ctid utterly deserved goal wigan congratulations worthy winners completely	Completely and utterly deserved goal for Wigan. Congratulations. Worthy winners on the day. #mcfc #CTID
FA Cup 2013 WAFC fans	ben watsoooonnnnnn manchester city 0 1	RT @InfoFootball_: BEN WATSOOOONNNNNN!!!!! Manchester City 0 - 1 Wigan
	win #wafc #facupfinal story delighted dave	Delighted for Dave Whelan. What a man, what a story, what a win! #WAFC #facupfinal
	lead #wafc deserved goal brilliant	GOAL! Brilliant. Deserved lead for #wafc

The high topic recall of Table 2 shows that our system is capable of detecting a variety of topics from live sports events, and is not limited to finding events of a pre-specified class. The good fit between team-specific tweets and team-related events shown in Figure 1 suggests that our classification of tweeters to fans is sufficiently accurate.

Note that after Wigan's late goal ('F' in Figure 1b), there is no spike in Wigan fans' tweets and the City fans become (and remain) almost silent. However, the examples in Table 3 show that our system did successfully identify the goal even without a spike, in contrast to the methods discussed earlier [10–12].

Table 3 gives examples of how different teams' fans discuss the same events in very different ways. Not only does this confirm our fan-team classification is effective, it also shows the potential power of generic topic detection. Journalists, and others seeking news, do not want just the headlines: they also want a variety of perspectives on a story. We have shown that by dividing active tweeters into sets, depending on which team they support, we can find two distinct views. For example, when Chelsea's goalkeeper, Petr Cech, narrowly prevented Liverpool equalizing, Chelsea fans reported this as a great save while Liverpool fans complained that the referee was mistaken and the ball had crossed the line. We believe our methods could be extended to cluster and analyse social media comments in other domains. For example, it may be possible to divide political commentators into groups depending on which party they support, allowing their varied views to be analysed separately, rather than mixed together.

Acknowledgments. This work is supported by the SocialSensor FP7 project, partially funded by the EC under contract number 287975.

References

1. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proc. of WWW 2010, pp. 851–860. ACM (2010)
2. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science 2(1), 1–8 (2011)
3. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting elections with Twitter: What 140 characters reveal about political sentiment. In: Proceedings of 4th ICWSM, vol. 10, pp. 178–185 (2010)
4. Newman, N.: The rise of social media and its impact on mainstream journalism. Reuters Institute for the Study of Journalism (2009)
5. FIFA TV: 2010 FIFA World Cup South Africa: Television audience report (2010)
6. Semiocast: Brazil becomes 2nd country on Twitter, Japan 3rd Netherlands most active country, Paris, France (January 2012)
7. Twitter: Twitter & TV: Use the power of television to grow your impact, `https://business.twitter.com/twitter-tv`
8. Goel, V., Stelter, B.: Social networks in a battle for the second screen. The New York Times (October 2, 2013)
9. Aiello, L., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., Goker, A., Kompatsiaris, I., Jaimes, A.: Sensing trending topics in Twitter. IEEE Transactions on Multimedia 15(6), 1268–1282 (2013)
10. van Oorschot, G., van Erp, M., Dijkshoorn, C.: Automatic extraction of soccer game events from Twitter. In: Proc. of the Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web (2012)
11. Zhao, S., Zhong, L., Wickramasuriya, J., Vasudevan, V.: Human as real-time sensors of social and physical events: A case study of Twitter and sports games. arXiv preprint arXiv:1106.4300 (2011)
12. Lanagan, J., Smeaton, A.F.: Using Twitter to detect and tag important events in live sports. In: Artificial Intelligence, pp. 542–545 (2011)