Mining Newsworthy Topics from Social Media

Carlos Martin, David Corney and Ayse Goker

Abstract Newsworthy stories are increasingly being shared through social networking platforms such as Twitter and Reddit, and journalists now use them to rapidly discover stories and eye-witness accounts. We present a technique that detects "bursts" of phrases on Twitter that is designed for a real-time topic-detection system. We describe a time-dependent variant of the classic *tf-idf* approach and group together bursty phrases that often appear in the same messages in order to identify emerging topics. We demonstrate our methods by analysing tweets corresponding to events drawn from the worlds of politics and sport, as well as more general mainstream news. We created a user-centred "ground truth" to evaluate our methods, based on mainstream media accounts of the events. This helps ensure our methods remain practical. We compare several clustering and topic ranking methods to discover the characteristics of news-related collections, and show that different strategies are needed to detect emerging topics within them. We show that our methods successfully detect a range of different topics for each event and can retrieve messages (for example, tweets) that represent each topic for the user.

1 Introduction

The growth of social networking sites, such as Twitter, Facebook and Reddit, is well documented. Every day, a huge variety of information on different topics is shared by many people. Given the real-time, global nature of these sites, they are used by many people as a primary source of news content [28]. Increasingly, such sites

C. Martin (✉) · D. Corney · A. Goker
IDEAS Research Institute, School of Computing & Digital Media
Robert Gordon University, Aberdeen AB10 7QB, Scotland, UK
e-mail: cmdanca@gmail.com

D. Corney
e-mail: dpacorney@gmail.com

A. Goker
e-mail: a.s.goker@rgu.ac.uk

M.M. Gaber et al. (eds.), *Advances in Social Media Analysis*,
Studies in Computational Intelligence 602,
DOI 10.1007/978-3-319-18458-6_2

are also used by journalists, partly to find and track breaking news but also to find user-generated content such as photos and videos, to enhance their stories. These often come from eye-witnesses who would be otherwise difficult to find, especially given the volume of content being shared.

Our overall goal is to produce a practical tool to help journalists and news readers to find newsworthy topics from message streams without being overwhelmed. Note that it is not our intention to re-create Twitter's own "trending topics" functionality. That is usually dominated by very high-level topics and memes, defined by just one or two words or a name and with no emphasis on 'news' nor any attempt to explain *why* something is trending.

The scale and diversity of these sites raise the question: how can users (whether journalists or non-professional "news consumers") find newsworthy topics from sites such as Twitter? One option would be for them to simply identify and follow some Twitter accounts that tend to Tweet regularly about the news, such as @CNN or @BBCNews. This approach has drawbacks however. Major news organizations tend to follow similar agendas, meaning that when an event occurs, either all the accounts will send equivalent messages, flooding the user with redundant messages, or none will send messages and the consumer will never learn of the story. The summer 2013 protests in Gezi Park, Turkey, were largely ignored by the Turkish national media for example and global mainstream media reports initially lagged behind social media reports. Such an approach will also miss what might be termed secondary messages from other sources, such as eye-witnesses, who may provide interesting and informative details about a story.

A similar problem occurs with a second possible option, namely using keywords (including hashtags) to filter incoming messages. This recasts the task as a search task with the attendant risks: all tweets that contain the search terms will be retrieved, including repetitions and redundant messages, while tweets that are relevant but do not contain the specified terms will be missed. A third option is to rely on Twitter's own "trending topics" algorithm, but, as noted above, this makes no attempt to filter for newsworthiness, and so tends to be dominated by celebrity news and Twitter memes.

Our system works by identifying phrases that show a sudden increase in frequency (a "burst") and then finding co-occurring groups of phrases to identify topics. Such bursts are typically responses to real-world events. In this way, the news consumer can avoid being overwhelmed by redundant messages, even if the initial stream is formed of diverse messages. The emphasis is on the temporal nature of message streams as we bring to the surface groups of messages that contain suddenly-popular phrases. An early version of this approach was recently described [2, 23], where it compared favourably to several alternatives and benchmarks. Here we expand and update that work, examining the effect of different clustering and topic ranking approaches used to form coherent topics from bursty phrases.

2 Related Work

Many of the individual techniques we use (as described in Sect. 3) have been used in related work, but not together and not in a user-centred way. No other study that we are aware of focuses on the needs of journalist and news-reading users, and with an emphasis on recency to augment traditional collection statistics such as *tf-idf*. It is our view that the domain must be tackled with a combination of such methods. We build on the idea of the importance of time and the concept of "sessions" now common in query-log analysis [18], but adapt it to the context of emerging news from Twitter.

Newman [27] discusses the central use of social media by news professionals, such as hosting live blogs of ongoing events. He also describes the growth of collaborative, networked journalism, where news professionals draw together a wide range of images, videos and text from social networks and provide a curation service. Broadcasters and newspapers can also use social media to increase brand loyalty across a fragmented media marketplace. Further examples of the use of live blogging by newspapers are given by Thurman et al. [41].

Schifferes et al. [38] discuss many of the issues around building a user-centred tool for professional journalists that identifies and verifies news from online social media. They discuss several examples of how information and misinformation has been spread rapidly through social media, showing both the potential benefits and risks of using Twitter as a news source. They interviewed a number of senior journalists, specialising in social media for mainstream news organizations, who expressed dissatisfaction with the tools currently available. Schifferes et al. suggest that independent measures of the reliability of contributors, content and context can help identify unreliable news and they describe a prototype verification system for automatic topic detection.

Petrovic et al. [32] focus on the task of first-story detection (FSD), which they also call "new event detection". They use a locality sensitive hashing technique on 160 million Twitter posts, hashing incoming tweet vectors into buckets in order to find the nearest neighbour and hence detect new events and track them. This work is extended in Petrovic et al. [33] using paraphrases for first story detection on 50 million tweets. Their FSD evaluation used newswire sources rather than Tweets, based on the existing TDT5 datasets. The Twitter-based evaluation was limited to calculating the average precision of their system, by getting two human annotators to label the output as being about an event or not. This contrasts with our goal here, which is to measure and improve the topic-level recall, i.e. to count how many newsworthy stories the system retrieved.

Benhardus [5] uses standard collection statistics such as *tf-idf*, unigrams and bigrams to detect trending topics. Two data collections are used, one from the Twitter API and the second being the Edinburgh Twitter corpus containing 97 million tweets, which was used as a baseline with some natural language processing used (e.g. detecting prepositions and conjunctions). The research focused on general trending topics (typically finding personalities and for new hashtags) rather than focusing the needs of journalistic users and news readers.

Shamma et al. [39] focus on "peaky topics" (topics that show highly localized, momentary interest) by using unigrams only. The focus of the method is to obtain peak terms for a given time slot when compared to the whole corpus rather than over a given time-frame. The use of the whole corpus favours batch-mode processing and is less suitable for real-time and user-centred analysis.

Phuvipadawat and Murata [34] analysed 154,000 tweets that contained the hashtag "#breakingnews". They determine popularity of messages by counting retweets and detecting popular terms such as nouns and verbs. This work is taken further with a simple *tf-idf* scheme that is used to identify similarity [35]; named entities are then identified using the Stanford Named Entity Recogniser in order to identify communities and similar message groups. Sayyadi et al. [37] also model the community to discover and detect events on the Live Labs SocialStream platform, extracting keywords, noun phrases and named entities. Ozdikis et al. [30] also detect events using hashtags by clustering them and finding semantic similarities between hashtags, the latter being more of a lexicographic method.

Ratkiewitcz et al. [36] focus specifically on the detection of a single type of topic, namely political abuse. Evidence used include the use of hashtags and mentions. Alvanaki [3] propose a system based on popular seed tags (tag pairs) which are then tracked, with any shifts detected and monitored. Becker et al. [4] also consider temporal issues by focusing on the online detection of real world events, distinguishing them from non-events (e.g. conversations between posters). Clustering and classification algorithms are used to achieve this. Methods such as *n*-grams and NLP are not considered. These methods do use natural language processing methods or *n*-grams, but many consider temporal factors in some way.

3 Methods

In this section we describe various aspects of our approach to topic detection and discuss how they work together. We consider "temporal document frequency-inverse document frequency" as a variation of the classic *tf-idf* to find trending terms at a specific point in time. We discuss several clustering methods to group these terms into topic-specific clusters and the use of *n*-grams to find phrases rather than isolated terms. We also consider the optimum speed with which to update results in real time, and compare methods to rank the results most usefully. In our experiments, we use collections of tweets (see Sect. 4.1), but the same approach should work for other streams of text messages.

3.1 BNgrams

Term frequency-inverse document frequency, or *tf-idf*, has been used for indexing documents since it was first introduced [40]. We are not interested in indexing documents however, but in finding novel trends, so we want to find terms that appear in one *time period* more than others. We treat temporal windows as documents and use them to detect words and phrases that are both new and significant. We therefore define newsworthiness as the combination of novelty and significance. We can maximise *significance* by filtering tweets either by keywords (as in this work) or by following a carefully chosen list of users, and maximise *novelty* by finding bursts of suddenly high-frequency words and phrases.

We select terms with a high "temporal document frequency-inverse document frequency", or $df - idf_t$, by comparing the most recent x messages with the previous x messages and count how many contain the term. We regard the most recent x messages as one "slot". After standard tokenization and stop-word removal, we index all the terms from these messages. For each term, we calculate the document frequency for a set of messages using df_{ti}, defined as the number of messages in a slot i that contain the term t.

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

This produces a list of terms which can be ranked by their $df - idf_t$ scores. Note that we add one to term counts to avoid problems with dividing by zero or taking the log of zero. To maintain some word order information, we define terms as *n*-grams, i.e. sequences of *n* words. Based on experiments reported elsewhere [23], we use 1-, 2- and 3-g in this work. High frequency *n*-grams are likely to represent semantically coherent phrases. Having found bursts of potentially newsworthy *n*-grams, we then group together *n*-grams that tend to appear in the same tweets. Each of these clusters defines a topic as a list of *n*-grams. We call this process of finding bursty *n*-grams "BNgrams."

3.2 Topic Clustering

An isolated word or phrase is often not very informative, but a group of them can define the essence of a story. Therefore, we group the most representative *n*-grams into clusters, each representing a single topic. A group of messages that discuss the same topic will tend to contain at least some of the same *n*-grams. We can then find the message that contains the most of these *n*-grams that define a topic, and use that message as the basis of a human-readable label for the topic. We now discuss three clustering algorithms that we compare here.

3.2.1 Hierarchical Clustering

Here, we initially assign every *n*-gram to its own singleton cluster, then follow a standard "group average" hierarchical clustering algorithm [26] to iteratively find and merge the closest pair of clusters. We define the similarity between two *n*-grams as the fraction of messages in the same slot that contain both of them, so it is highly likely that the term clusters whose similarities are high represent the same topic.

The clustering is repeated until the similarity between the nearest un-merged clusters falls below a fixed threshold θ, producing the final set of topic clusters for a set of tweets. In our experiments, we use a similarity threshold of $\theta = 0.5$ which means that two terms must appear in at least half of the same tweets in order to belong to the same topic. Note that this threshold implicitly defines the number of clusters that the system returns for any given set of tweets. If we use a very low threshold, then we will merge clusters that only share a few terms, which will tend to lead to a very small number of large clusters. A high threshold will conversely lead to a very large number of small clusters with very few overlapping terms. This gives a potential means to control the granularity of detected topics. Preliminary results suggest that the exact value is not critical.

Individual messages are then assigned to the cluster that they share the most terms with, if any. Note that not every tweet will be assigned to any topic. This is deliberate, as many tweets are not newsworthy and/or do not fall into the same topic category as any other tweets.

Further details about this algorithm and its parameters can be found in our previous published work [2].

3.2.2 Apriori Algorithm

The Apriori algorithm [1] finds all the associations between the most representative *n*-grams based on the number of tweets in which they co-occur. Each association is a candidate topic at the end of the process. One of the advantages of this approach is that one *n*-gram can belong to different associations (i.e. it allows partial membership), avoiding one problem with hierarchical clustering. The number of associations does not have to be specified in advance. We also obtain maximal associations after clustering to avoid large overlaps in the final set of topic clusters.

One parameter associated to this technique is the *support value* which determines the minimum number of documents a group of *n*-grams (association) should share to be considered as a candidate topic. The value of this parameter represents a percentage of all the documents from the corresponding slot. Preliminary experiments considering different values of this parameter suggested we fix its value to 0. It means that no candidate topic is discarded. In addition, maximal associations are obtained at the end of the approach to avoid overlaps in the final candidate topics set. The main idea of this approach is to delete all the associations whose keywords are contained in another association and sharing most of the topic tweets with the previous one. This second requirement was introduced to confirm that both topics are talking about the same matter before they are merged into a single topic.

3.2.3 Gaussian Mixture Models (GMM)

GMMs assign probabilities (or strengths) of membership of each *n*-gram to each cluster, allowing partial membership of multiple clusters. This approach does require the number of clusters to be specified in advance, although this can be automated (e.g. by using Bayesian information criteria [15]). Here, we use the Expectation—Maximisation algorithm to optimise a Gaussian mixture model [12]. We fix the number of clusters at 20, although initial experiments showed that using more or fewer produced very similar results. A more sophisticated variation would be to vary this value as a function of the number of messages in the slot. Seeking more clusters in the data than there are newsworthy topics means that some clusters will contain irrelevant tweets and outliers, which can later be assigned a low rank and effectively ignored, leaving us with a few highly-ranked clusters that are typically newsworthy.

We use the Weka implementation [17], which iteratively fits spherical Gaussian components to the data.

3.3 Topic Ranking

To maximise usability we need to avoid overwhelming the user with a very large number of topics. We therefore want to rank (and potentially filter) the results by relevance, in the same fashion as typical search engines. Here, we compare two topic ranking techniques.

3.3.1 Maximum *n*-gram $df - idf_t$

One method is to rank topics according to the maximum $df - idf_t$ value of their constituent *n*-grams. The motivation of this approach is the assumption that the most popular *n*-gram from each topic represents the core of the topic.

3.3.2 Weighted Topic-Length

As an alternative we propose weighting the topic-length (i.e. the number of terms found in the topic) by the number of tweets in the topic to produce a score for each topic. Thus the most detailed and popular topics are assigned higher rankings. The use of clustering techniques that allow each *n*-gram to have partial membership of different clusters suggests the need for an alternative topic ranking technique, because the previous method may fail to give a good performance if the top-*m* results from the ranking have several and diverse topics at the same time. We define this score thus:

$$s_t = \alpha \cdot \frac{L_t}{L_{max}} + (1 - \alpha) \cdot \frac{N_t}{N_s} \tag{2}$$

where s_t is the score of topic t, L_t is the length of the topic, L_{max} is the maximum number of terms in any current topic, N_t is the number of tweets in topic t and N_s is the number of tweets in the slot. Finally, α is a weighting term. Setting α to 1 rewards topics with more terms; setting α to 0 rewards topics with more tweets. We used $\alpha = 0.7$ in our experiments, giving slightly more weight to those stories containing more details, although the exact value is not critical.

4 Experiments

Here, we show the results of our experiments with several variations of the BNgram approach. We focus on two questions. First, what is best slot size to balance topic recall and refresh rate? A very small slot size might lead to missed stories as too few tweets would be analysed; conversely, a very large slot size means that topics would only be discovered some time after they have happened. This low 'refresh rate' would reduce the timeliness of the results. Second, what is the best combination of clustering and topic ranking techniques? In Sect. 3, we introduced three clustering methods and two topic ranking methods; we need to determine which methods are most useful.

We have previously shown that our methods perform well [2]. The BNgram approach was compared to a popular baseline system in topic detection and tracking—Latent Dirichlet Allocation (LDA) [6]—and to several other competitive topic detection techniques, getting the best overall topic recall. In addition, we have shown the benefits of using *n*-grams compared with single words for this sort of analysis [23]. Below, we present and discuss the results from our current experiments, starting with our approach to evaluation.

4.1 Evaluation Methods

When evaluating any information retrieval system, it is crucial to define a realistic test problem. We used three Twitter data sets focused on popular real-world events and compare the topics that our algorithm finds with an externally-defined ground truth. To establish this ground truth, we relied on mainstream media (MSM) reports of the three events. This use of MSM sources helps to ensure that our ground truth topics are newsworthy (by definition) and that the evaluation is goal-focussed (i.e. will help journalists write such stories). We see no reason why our methods would not work on non-MSM stories, if they are discussed on the online social networks. However, this is harder to evaluate given the lack of a convenient ground-truth.

We filtered Twitter using relevant keywords and hashtags to collect tweets around three events: the 2012 "Super Tuesday" primaries, part of the presidential nomination race of the US Republican Party; the 2012 FA Cup final, the climax to the English football season; and the 2012 US presidential election, an event of global significance.

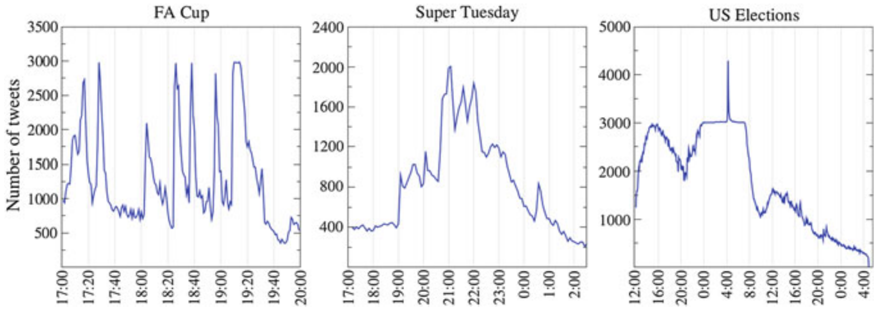

Fig. 1 Twitter activity during events (tweets per minute). For the FA Cup, the peaks correspond to start and end of the match and the goals. For the two political collections, the peaks correspond to the main result announcements

In each case, we reviewed the published MSM accounts of the events and chose a set of stories that were significant, time-specific, and represented on Twitter. For example, we ignored general reviews of the state of US politics (not time-specific), and quotes from members of the public (not significant events).

Using MSM sources presents its own problems however. Each MSM source has its own policy for selecting and sharing stories, most obviously being national biases (e.g. UK outlets tend to emphasise UK stories). To get detailed accounts of the events of interest, we relied on "live blogs" produced by various MSM outlets [41]; again, many events are not covered by live blogs, potentially introducing a further bias into our selection of topics. Of course the choice of *which* MSM sources to use is critical and to some extent subjective. We chose MSM outlets that have an excellent reputation for timely and reliable reporting, primarily the BBC and the Wall Street Journal.

For each target topic, we identified around 5–7 keywords that defined the story and used these to measure recall and precision, as discussed below. Some examples are shown in the first two columns of Table 4. We also defined several "forbidden" keywords. A topic was only considered as successfully recalled if all of the "mandatory" terms were retrieved and *none* of the "forbidden" terms. The aim was to avoid producing topics such as "victory Romney Paul Santorum Gingrich Alaska Georgia" that convey no information about who won or where; or "Gingrich wins", which is too limited to define the story because it doesn't name the state where the victory occurred. Similarly, when detecting events during a football match, topics labels such as "Liverpool Chelsea goal" or just "goal" are not useful.

Figure 1 shows the frequency of tweets collected over time, with further details in Ref. [2]. We have made all the data freely available, including the ground truth topics.[1] The three data sets differ in the rates of tweets, determined by the popularity of the topic and the choice of filter keywords. The mean tweets per minute (tpm) were: Super Tuesday, 832 tpm; FA Cup, 1293 tpm; and US elections, 2209 tpm. For a slot size of 1500 tweets these correspond to a "topic refresh rate" of 108, 70 and

[1] http://www.socialsensor.eu/results/datasets/72-twitter-tdt-dataset.

41 s respectively. This means that a user interface displaying these topics could be updated every 1–2 min to show the current top-10 (or top-m) stories.

To generate the sets of Tweets used in the evaluation, we crawled Twitter during the events using appropriate sets of filter keywords, such as the names of the participants. For timetabled events, such as elections and sports fixtures, such keywords are easy to define in advance and help to ensure that the topics discovered are newsworthy. For less predictable breaking news stories, such as natural disasters, other approaches may be more appropriate. For example, a list of reliable, news-related Twitter accounts can be created and their Tweets analysed; this is the subject of ongoing work. Even such straightforward approaches to filtering help to mitigate the fact that many bursty topics on Twitter would not usually be considered newsworthy [8].

We ran the topic detection algorithm on each data set. This produced a ranked list of topics, each defined by a set of terms (i.e. n-grams). For our evaluation, we focus on the recall of the top m topics ($1 \leq m \leq 10$) at the time each ground-truth story emerges. For example, if a particular story was being discussed in the mainstream media from 10:00–10:15, then we consider the topic to be recalled if the system ranked it in the top m at any time during that period.

The automatically detected topics were compared to the ground truth (comprising 22 topics for Super Tuesday; 13 topics for FA Cup final; and 64 topics for US elections) using three metrics:

- Topic recall: Percentage of ground truth topics that were successfully detected. A topic was considered successfully detected if the automatically produced set of words contained all mandatory keywords for it (and none of the forbidden terms, if defined).
- Keyword precision: Percentage of correctly detected keywords out of the total number of keywords for all topics detected by the algorithm in the slot.
- Keyword recall: Percentage of correctly detected keywords divided by the total number of ground truth keywords (excluding forbidden keywords) in the slot. One key difference between "topic recall" and "keyword recall" is that the former is a user-centred evaluation metric, as it considers the power of the system at retrieving and displaying to the user stories that are meaningful and coherent, as opposed to retrieving only some keywords that are potentially meaningless in isolation.

Note that we do not attempt to measure topic precision as this would need an estimate of the total number of newsworthy topics at any given time, in order to verify which (and how many) of the topics returned by our system were in fact newsworthy. This would require an exhaustive manual analysis of MSM sources to identify every possible topic (or some arbitrary subset), which is infeasible. One option is to compare detected events to some other source, such as Wikipedia, to verify the significance of the event [29], but Wikipedia does not necessarily correspond to particular journalists' requirements regarding newsworthiness and does not claim to be complete. The scores reported below were automatically computed by an evaluation script. However, to ensure the reliability of results, we conducted several rounds of manual evaluation of results and confirmed their agreement with the automatically produced ones.

Table 1 Topic recall for different slot sizes (with hierarchical clustering)

Slot size (tweets)	500	1000	1500	2000	2500
Super tuesday	**0.773**	0.727	0.682	0.545	0.682
FA cup	0.846	0.846	**0.923**	**0.923**	**0.923**
US elections	0.750	0.781	**0.844**	0.734	0.766
Weighted mean	0.77	0.78	**0.82**	0.72	0.77

4.2 Results

Table 1 shows the effect on topic recall when varying the slot size, with the same total number of topics in the evaluation for each slot size. The mean is weighted by the number of topics in the ground truth for each set, giving greater importance to larger test sets. Overall, using very few tweets produces slightly worse results than with larger slot sizes (e.g. 1500 tweets), presumably as there is too little information in such a small collection. Slightly better results for the Super Tuesday set occur with fewer tweets; this could be due to the slower tweet rate in this set. Note that previous experiments [23] showed that including 3-g improves recall compared to just using 1- and 2-g, but adding 4-g provides no extra benefit, so here we use 1-, 2- and 3-g phrases throughout.

Lastly, we compared the results of combining different clustering techniques with different topic ranking techniques (see Fig. 2). We conclude that the hierarchical clustering performs well despite the weakness discussed above (i.e. each n-gram is assigned to only one cluster), especially in FA Cup dataset. Also, the use of weighted topic-length ranking technique improves topic recall with hierarchical clustering in the political data sets.

The Apriori algorithm performs quite well in combination with the weighted topic length ranking technique (note that this ranking technique was specially created for the "partial" membership clustering techniques). We see that the Apriori algorithm in combination with the maximum n-gram $df - idf_t$ ranking technique produces slightly worse results, as this ranking technique does not produce diverse topics for the first results (from top 1 to top 10, in our case) as we mentioned earlier.

Turning to the EM Gaussian mixture model results, we see that this method works very well on the FA Cup final and US elections data sets. Despite being a "partial" membership clustering technique, the use of weighted topic length ranking technique does not make any representative difference, even its performance is worse in Super Tuesday dataset. Further work is needed to test this.

Table 2 summarises the results of the three clustering methods and the two ranking methods across all three data sets. The weighted-mean scores show that for the three clustering methods, ranking by the length of the topic is more effective than ranking by each topic's highest $df - idf_t$ score. We can see that for the FA Cup set, the Hierarchical and GMM clustering methods are the best ones in combination with

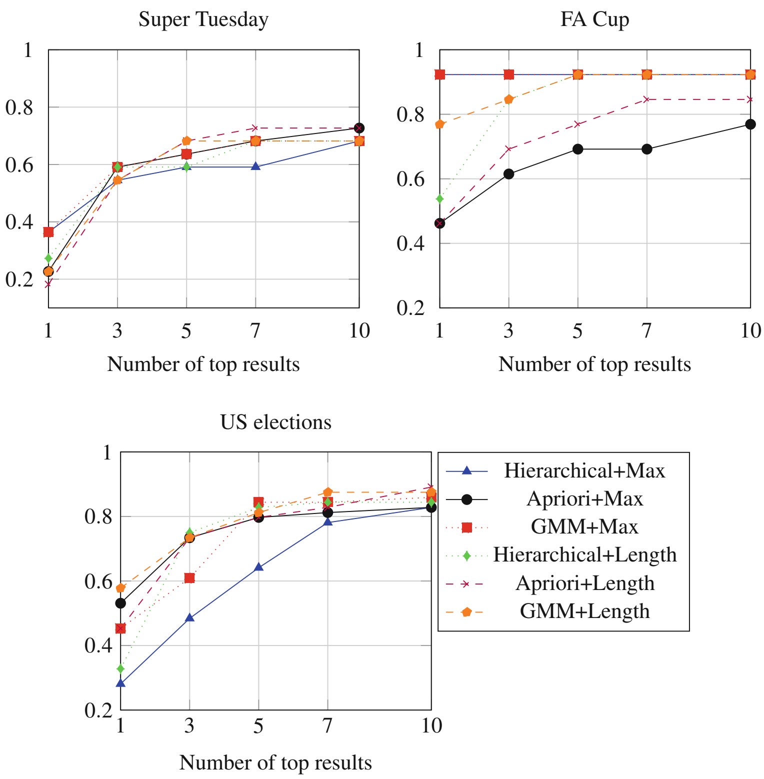

Fig. 2 Topic recall for different clustering techniques in the Super Tuesday, FA Cup and US elections sets (slot size = 1500 tweets)

Table 2 Normalised area under the curve for the three datasets combining the different clustering and topic ranking techniques (1500 tweets per slot)

Ranking	Max. n-gram $df - idf_t$			Weighted topic-length		
Clustering	Hierar.	Apriori	GMM	Hierar.	Apriori	GMM
FA cup	**0.923**	0.677	**0.923**	0.861	0.754	0.892
Super tuesday	0.573	0.605	0.6	0.591	**0.614**	0.586
US elections	0.627	0.761	0.744	0.761	0.772	**0.797**
Weighted mean	0.654	0.715	0.735	0.736	0.734	**0.763**

the maximum n-gram $df - idf_t$ ranking technique. For Super Tuesday and US Elections data sets, "partial" membership clustering techniques (Apriori and GMM, respectively) perform the best in combination with weighted topic length ranking technique, as expected.

Finally, Table 3 shows more detailed results, including keyword precision and recall, for the best combinations of clustering and topic ranking methods of the three datasets when the top five results are considered per slot. In addition, Table 4 shows some examples of ground truth and BNgram detected topics and tweets within the corresponding detected topics for all datasets.

Table 3 Best results for the different datasets after evaluating top 5 topics per slot

Method	*T-REC@5*	*K-PREC@5*	*K-REC@5*
Super tuesday			
Apriori+Length	0.682	0.431	0.68
GMM+Length	0.682	0.327	0.753
FA cup			
Hierar.+Max	0.923	0.337	0.582
Hierar.+Length	0.923	0.317	0.582
GMM+Max	0.923	0.267	0.582
GMM+Length	0.923	0.162	0.673
US elections			
GMM+Max	0.844	0.232	0.571

T-REC, K-PREC, and K-REC refers to topic-recall and keyword-precision/recall respectively

Table 4 Examples of the mainstream media topics, the target keywords, the topics extracted by the $df - idf_t$ algorithm, and example tweets selected by our system from the collections

Target topic	Ground truth keywords	Extracted keywords	Example tweet
Newt Gingrich says "Thank you Georgia! It is gratifying to win my home state so decisively to launch our March Momentum"	Newt Gingrich, Thank you, Georgia, March, Momentum, gratifying	launch, March, Momentum, decisively, thank, Georgia, gratifying, win, home, state, #MarchMo, #250gas, @newtgingrich	@**Bailey_Shel**: RT @newtgingrich: Thank you Georgia! It is gratifying to win my home state so decisively to launch our March Momentum. #MarchMo #250gas
Salomon Kalou has an effort at goal from outside the area which goes wide right of the goal	Salomon Kalou, run, box, mazy	Liverpool, defence, before, gets, ambushed, Kalou, box, mazy, run, @chelseafc, great, #cfcwembley, #facup, shoot	@**SharkbaitHooHa_**: RT @chelseafc: Great mazy run by Kalou into the box but he gets ambushed by the Liverpool defence before he can shoot #CFCWembley #FACup
US President Barack Obama has pledged "the best is yet to come", following a decisive re-election victory over Republican challenger Mitt Romney	Obama, best, come	America, best, come, United, States, hearts, #Obama, speech, know, victory	@**northoaklandnow**: "We know in our hearts that for the United States of America, the best is yet to come," says #Obama in victory speech

5 Applications

In this section, we discuss several specific applications of our clustered BNgram approach. These go some way to demonstrate the robustness of the algorithm and explore how it can be applied to sports, subjective event summarization and rolling 24-h news.

5.1 *Finding Events in Football Matches*

Earlier in this chapter (Sect. 4), we described analysis of the 2012 FA Cup Final. In a recent paper [9], we also compared this with the 2013 Final, a match between Manchester City and Wigan Athletic. We used this as a further means to evaluate the automated topic detection system, once again using a ground truth derived from mainstream media.

In this case, as well as detecting topics, we also attempted to identify the team that each Twitter user supported, or to recognise their neutrality. For each user, we counted the number of times they mentioned each team in all their tweets. An initial manual inspection showed that fans tend to use their team's standard abbreviation as a hashtag (e.g. #CFC or #MCFC) a great deal 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 "fan" any user with a degree of two or more and treat everyone else as neutral. A manual evaluation indicated that this approach identifies which team is supported for over 90 % of the tweeters but occasionally mis-identifies neutral reporters as supporting one team. We believe this could be improved if we extended the analysis over several matches to build up more evidence for support.

We chose a total of 25 events from the two matches that were reported by the mainstream media (specifically the BBC commentaries). Of these our system found around 75–90 % of the events. The variation is likely due to the different nature of the two matches considered and the volume of tweets generated. Repeating over more matches would give us a clearer indication of quality, but clearly the algorithm does find a large number of the most important events.

Other systems have been proposed to discover events within sporting fixtures, but these typically are designed to find only events of pre-defined classes, such as goals or bookings [42, 43]. In contrast, ours is agnostic about the specific nature of the event, relying only on shifts in the word-use used by multiple users to describe it.

We also showed that fans of each team tended to give biased, subjective views of the events, as would be expected. We explored this further in our next paper (see Sect. 5.2).

5.2 *Subjective Summarization of Sporting Events*

In most journalism, there is an aim of objectively summarizing the events and presenting them from a neutral point of view, although there is some debate about how much this is really possible, or even desirable [11]. However, when producing such a neutral point of view, there is a risk that the distinct opinions expressed by different groups get lost in the mix. This range of opinions is often of interest to journalists and to news consumers, as it reflects diversity. In a democracy, it is important that different arguments are presented and considered, and this may effect people's opinions. Related work on automated document summarization has sometimes attempted to distinguish and summarize diverse opionions [20], but this is rare. In sports, fans may rarely change their allegiance to one team or another, but it is still interesting to consider the range of opinions expressed. Our work here can be seen as a step towards subjective event summarization, summarizing messages from specific, distinct points-of-view.

Having shown that we could (a) identify key events of football matches, and (b) identify which team each tweeter supported, we then combined these two methods into a subjective event summarization tool, as described in a recent paper [10]. We used the same method as before to estimate which team each tweeter supports, if any (Sect. 5.1), and the same BNgram topic detection methods. This time, we also tracked the relatively objective, neutral mainstream media comments from the BBC's live text-based commentaries. For each slot, we used our BNgram algorithm to select up to 10 topics. We then compared these to the corresponding BBC commentary, using a simple cosine similarity, and selected the most similar. In this way, we could discover what each set of fans were *subjectively* saying about the events that were *objectively* most important. As an alternative to the BBC commentary, we could have used the BNgram algorithm on the entire collection of tweets, thus incorporating the view of both sets of fans and the many neutral observers, when determining the "objective" event list.

A distinct but related approach is to identify reliable "reporters" of events, such as people watching a football match who also provide regular, accurate tweets about it [21]. In common with several event-detection approaches [42, 43], they rely on spikes in the overall activity of message streams to identify events, unlike our work which only needs the frequency of terms to shift within a (potentially) unchanging volume of messages.

Although not strictly related to topic detection, we have also analysed tweets sent by fans of different teams during English Premier League matches. In that work [7], we focussed on the use of swearing in tweets and how curse words are used to express sentiment, both positive and negative. This contrasts with an assumption common to much sentiment analysis research, that swearing is more typically negative or sarcastic, and rarely positive [22, 25].

5.3 Real-Time Topic Detection for 24 Hours of News

Our previous studies described above have all focussed on specific, pre-specified events: the Super Tuesday primaries; the 2012 US Presidential Elections; and the 2012 and 2013 FA Cup Finals. While very useful as benchmarks, there is a risk that methods developed to analyse such specific events may fail to generalise to the wider case of finding newsworthy stories during a typical 24-h news cycle. To test our approach in this scenario, we entered the 2014 Social News On the Web (SNOW) Data Challenge[2] [24]. This challenge is held in conjunction with the 23rd International World Wide Web Conference (WWW 2014).

The task of this challenge is to retrieve newsworthy stories or topics for multiple timeslots over 24 h, where each timeslot is 15 min long. The required format of each topic includes a human-readable label, a set of the most representative keywords, a set of tweets that are related to the story and links to any relevant images from the tweets. The tweets identified for each event could be from the corresponding timeslot or any earlier one (but not later), to simulate a real-time scenario.

As the guidelines of the challenge show, the extracted topics were evaluated on several dimensions, namely: precision and recall, readability, coherence, relevance and diversity. Further details can be found in the official description of this challenge [31].

Regarding our BNgram approach, we modified the strategy slightly to select bursty terms after analysis of topics produced during previous experiments. Bigrams, trigrams, entities, hashtags and URLs were considered as terms in the SNOW experiments. The Apriori algorithm (see Sect. 3.2) was used for the clustering algorithm. In addition, we considered temporal windows (i.e. timeslots) instead of using a fixed number of tweets per slot, and used two previous timeslots for the penalization of common terms. The final formula to compute $df - idf$ scores was (mostly based on Eq. 1):

$$df - idf_{ti} = (df_{ti} + 1) \cdot \frac{1}{\log\left(\frac{\sum_{j=i}^{s} df_{t(i-j)}}{s} + 1\right) + 1}. \qquad (3)$$

where $s = 2$ in the experiments as before.

To populate topics with tweets, our approach creates a query based on the most representative terms to retrieve the associated tweets to the story. In addition, replies to the previous tweets are also considered as they can add further details of the story. The main reasons to include them is that they are not text-query dependant and add a wider range of people's view in many cases. However, we believe a filtering process should be considered for these replies, as we detected many spam replies, such as advertising links.

Our topic label approach here is based on the selection of the most representative tweet from the set of topic tweets, following some recent advice that headlines

---

[2] http://www.snow-workshop.org/.

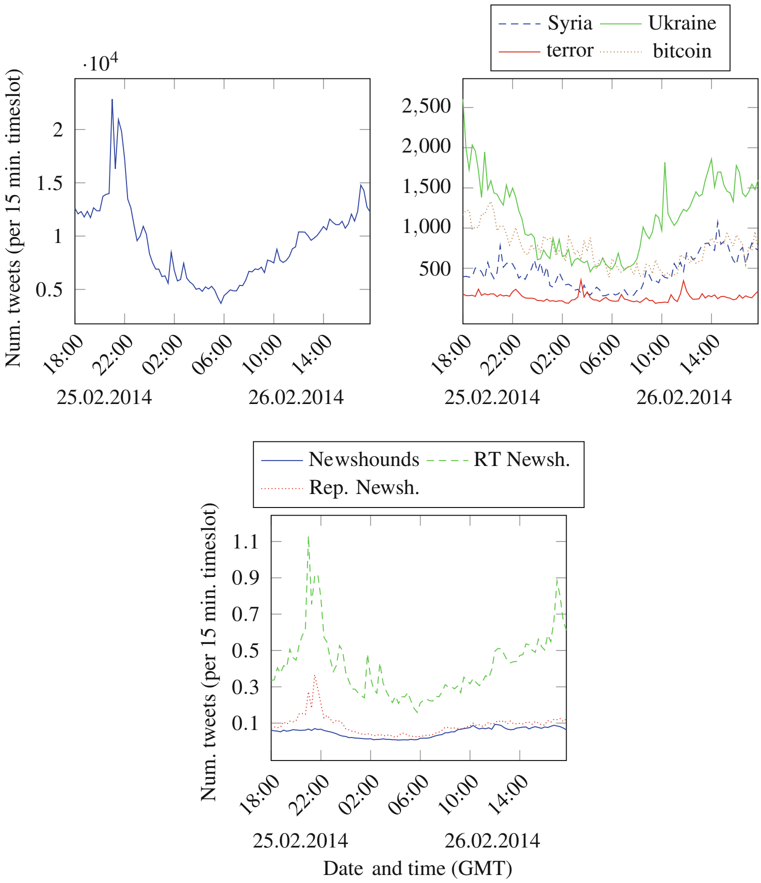

Fig. 3 Analysis of data test collection

increasingly resemble tweets.[3] Therefore, the tweet containing the greatest number of topic terms and duplicates (e.g. retweets) is selected. Its text, is then "cleaned" (by removing redundant user mentions, URLs and abbreviations such as RT and MT) to make it more readable, and is then used as the topic title or label.

The final topics are ranked by their bursty scores where each score is the maximum $df - idf_t$ value of their constituent terms (see Sect. 3.3). Our assumption is that the the most popular term from each topic represents the core of the topic and diverse topics are detected by the algorithm as the collection is not event-based.

Our final test data collection was composed of 901,895 tweets and stored in Solr after filtering out the non-English tweets. We extracted entities from each tweet using the Stanford NLP library [14], and created links between replies and retweets with their original tweets.

[3] http://perryhewitt.com/5-lessons-buzzfeed-harvard/.

Table 5 Examples of topics about the tracking keywords

Timeslot	Topic label	Keywords	Tweets
Syria			
25/2/14 20:30	Al Qaeda branch in Syria issues ultimatum to splinter group: The head of an al Qaeda-inspired militia fighting	Militia, fighting, branch, issues, ultimatum, splinter, group, inspired, head, syria, al qaeda	Al Qaeda branch in Syria issues ultimatum to splinter group http://t.co/gQDm0p7Wur
			Al Qaeda ultimatum to splinter group: The head of an al Qaeda-inspired militia fighting in Syria is giving a... http://t.co/9KFu1CG1F6
26/2/14 00:15	Jordan Bahrain Morocco Syria Qatar Oman Iraq Egypt United States 346	346	Jordan Bahrain Morocco Syria Qatar Oman Iraq Egypt United States 346 http://t.co/RjZAwwMJ95
Terror			
26/2/14 4:00	25 marines to arrest 'worlds biggest drug lord' El Chapo Guzman 73 anti-terror-squad police to arrest 'Internet entrepreneur' Kim Dotcom	Arrest, worlds, biggest, 25, marines, terror, squad, amp, police, 73, anti, drug, lord, internet, entrepren,el chapo guzman	25 marines to arrest 'worlds biggest drug lord' El Chapo Guzman 73 anti-terror-squad & police to arrest 'Internet entrepreneur' Kim Dotcom
Ukraine			
26/2/14 10:15	Ukraine minister disbands Berkut riot police blamed for violence—CNN	Riot, police, disbands, blamed, violence, ukraine, cnn	RT @BBCWorld: Ukraine disbands elite Berkut anti-riot police unit, acting interior minister says http://t.co/5GqM6jjryu
			RT @cnnbrk: Ukraine has disbanded a riot police force used against anti-government protesters, acting interior minister said
			RT @BBCGavinHewitt: In Ukraine the Berkut special police units blamed for most of the shootings have been disbanded

(continued)

Table 5 (continued)

Timeslot	Topic label	Keywords	Tweets
Bitcoin			
25/2/14 20:00	Mt. Goxs Demise Marks The End of Bitcoins First Wave Of Entrepreneurs	Demise, marks, end, first, wave, gox, bitcoin, entrepreneurs	Mt. Gox's Demise Marks The End of Bitcoin's First Wave Of Entrepreneurs http://t.co/gIKKP3RLQn by @kimmaicutler
			Mt. Gox's Demise Marks The End of Bitcoin's First Wave Of... http://t.co/X7iUKN3Vsv #eCommerce #Finance #Startups #TC #techcrunch #tech
			#SuryaRay #Surya #SuryaRay #Surya Mt. Gox's Demise Marks The End of Bitcoin's... http://t.co/csX2dB26w4 @suryaray @suryaray @suryaray3

Figure 3a shows the distribution of tweets per 15-min timeslot for this final collection going from 18:00 25/2/2014 to 18:00 26/2/2014. Note that the high peak shown from 20:00 to 22:00 on 25/2/2014 corresponds to tweets (mainly retweets and replies, as shown in Fig. 3c) related to several Champions League football matches that were taking place at that time. This is because there were some sport commentators in the list of accounts used to select the tweets (as provided by the SNOW challenge organizers). There are no clear peaks during that period related to the keywords provided for tracking, as shown by Fig. 3b, as these are not football related. Finally, the activity goes down overnight as most of the Twitter accounts being followed are UK-based, so it is more likely they were inactive during these hours.

Table 5 shows some representative topics associated to the tracked keywords, giving some indication of the quality of the stories found.

The official evaluation results of our method in the SNOW Data Challenge are included in Papadopoulos et al. [31]. Overall, our submission was placed second out of the eleven teams from round the world that completed the challenge. The winning team of Ifrim et al. also used our BNgram approach to rank and filter topics, alongside more aggressive pre-processing and filtering methods [19]. While neither team found every one of the target topics defined by the challenge organizers, the fact that the two best-placed teams used variations of the same BNgram algorithm strongly suggests that this is a robust and flexible tool for detecting topics in Twitter streams.

6 Conclusions

In Sect. 4, we presented our main findings regarding the power of the BNgram algorithm. If we compare the results between the three main collections, one difference is particularly striking: the topic recall is far higher for football (over 90 %) than for politics (around 60–80 %; Table 2). This is likely to reflect the different nature of conversations about the events. Topics within a live sports event tend to be transient: fans care (or at least tweet) little about what happened 5 min ago; what matters is what is happening "now". This is especially true during key events, such as goals, as also discussed in Sects. 5.1 and 5.2. In politics, conversations and comments tend to spread over hours (or even days) rather than minutes. This means that sports-related topics tend to occur over a much narrower window, with less overlapping chatter. In politics, several different topics are likely to be discussed at the same time, making this type of trend detection much harder. Looking back at the distribution of the tweets over time (Fig. 1), we can see clear spikes in the FA Cup graph, each corresponding to a major event (kick-off, goals, half-time, full-time etc.). No such clarity is in the politics graphs, which instead is best viewed as many overlapping trends.

This difference is reflected in the way that major news stories often emerge: an initial single, focussed story emerges but is later replaced with several potentially overlapping sub-stories covering different aspects of the story. Our results suggest that a dynamic approach may be required for newsworthy topic detection, finding an initial clear burst and subsequently seeking more subtle and overlapping topics. The specific applications we described included analysis of 24 h of news-related tweets (Sect. 5.3). In this work, we saw more clearly that news stories tend to emerge over time, to overlap greatly and to have multiple angles. As more details emerge around breaking news stories, it becomes increasingly important to go further than topic detection and to start identifying links between topics.

Recently, Twitter has been actively increasing its ties to television.[4] Broadcast television and sporting events share several common features: they occur a pre-specified times; they attract large audiences; and they are fast-paced. These features all allow and encourage audience participation in the form of sharing comments and holding discussions during the events themselves, such that the focus of the discussion is constantly moving with the event itself. Potentially, this can allow targeted time-sensitive promotions and advertising based on topics currently receiving the most attention. Facebook and other social media are also competing for access to this potentially valuable "second screen" [16]. Television shows are increasingly promoting hashtags in advance, which may make collecting relevant tweets more straightforward. One potential approach to help with this is a "visual backchannel" [13] that allows users to visualize and make sense of masses of streaming information, and this does could be enhanced with improved topic detection and clustering.

[4] "Twitter & TV: Use the power of television to grow your impact" https://business.twitter.com/twitter-tv.

Even if topic detection for news requires slightly different methods or parameters when compared to detecting sporting and live television events, all these areas have substantial and growing demand.

Acknowledgments This work is supported by the SocialSensor FP7 project, partially funded by the EC under contract number 287975. We wish to thank Nic Newman and Steve Schifferes of the Department of Journalism, City University London and Andrew MacFarlane of the Department of Computer Science, City University London, for their invaluable advice.

References

1. Agrawal, R., Srikant, R., et al.: Fast algorithms for mining association rules. In: Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, vol. 1215, pp. 487–499 (1994)
2. Aiello, L., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., Göker, A., Kompatsiaris, I., Jaimes, A.: Sensing trending topics in Twitter. IEEE Trans. Multimedia **15**(6), 1268–1282 (2013). doi:10.1109/TMM.2013.2265080
3. Alvanaki, F., Sebastian, M., Ramamritham, K., Weikum, G.: Enblogue: emergent topic detection in Web 2.0 streams. In: Proceedings of the 2011 International Conference on Management of Data, pp. 1271–1274. ACM (2011)
4. Becker, H., Naaman, M., Gravano, L.: Beyond trending topics: real-world event identification on Twitter. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM11) (2011)
5. Benhardus, J.: Streaming trend detection in Twitter. National Science Foundation REU for Artificial Intelligence, Natural Language Processing and Information Retrieval, University of Colarado (2010)
6. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
7. Byrne, E., Corney, D.: Sweet FA: sentiment, swearing and soccer. In: ICMR2014 1st Workshop on Social Multimedia and Storytelling. Glasgow, UK (2014)
8. Castillo, C., Mendoza, M., Poblete, B.: Information credibility on Twitter. In: Proceedings of the 20th International Conference on World Wide Web, pp. 675–684. ACM (2011)
9. Corney, D., Martin, C., Göker, A.: Spot the ball: detecting sports events on Twitter. In: ECIR 2014, pp. 449–454. Amsterdam, Holland (2014)
10. Corney, D., Martin, C., Göker, A.: Two sides to every story: Subjective event summarization of sports events using Twitter. In: ICMR2014 1st Workshop on Social Multimedia and Storytelling. Glasgow, UK (2014)
11. Cunningham, B.: Re-thinking objectivity. Columbia. Journalism Rev. **42**(2), 24–32 (2003)
12. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B (Methodological), 1–38 (1977)
13. Dork, M., Gruen, D., Williamson, C., Carpendale, S.: A visual backchannel for large-scale events. IEEE Trans. Vis. Comput. Graph. **16**(6), 1129–1138 (2010)
14. Finkel, J.R., Grenager, T., Manning, C.: Incorporating non-local information into information extraction systems by Gibbs sampling. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL '05, pp. 363–370. Stroudsburg, PA, USA (2005). doi:10.3115/1219840.1219885
15. Fraley, C., Raftery, A.E.: How many clusters? Which clustering method? Answers via model-based cluster analysis. Comput. J. **41**(8), 578–588 (1998)
16. Goel, V., Stelter, B.: Social networks in a battle for the second screen. The New York Times (2013). http://www.nytimes.com/2013/10/03/technology/social-networks-in-a-battle-for-the-second-screen.html. Accessed 24 Mar 2014

17. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. ACM SIGKDD Explor. Newsl. **11**(1), 10–18 (2009)
18. He, D., Göker, A., Harper, D.: Combining evidence for automatic web session identification. Inf. Process. Manage. **38**(5), 727–742 (2002)
19. Ifrim, G., Shi, B., Brigadir, I.: Event detection in Twitter using aggressive filtering and hierarchical tweet clustering. In: Proceedings of the SNOW 2014 Data Challenge (2014)
20. Ku, L.W., Lee, L.Y., Wu, T.H., Chen, H.H.: Major topic detection and its application to opinion summarization. In: 28th ACM SIGIR Conference, pp. 627–628. ACM (2005)
21. Kubo, M., Sasano, R., Takamura, H., Okumura, M.: Generating live sports updates from Twitter by finding good reporters. In: IEEE/WIC/ACM International Joint WI-IAT Conferences, vol. 1, pp. 527–534. IEEE (2013)
22. Liu, B.: Sentiment analysis and subjectivity. In: N. Indurkhya, F.J. Damerau (eds.) Handbook of Natural Language Processing, 2nd edn. Chapman & Hall, Boca Raton (2010)
23. Martin, C., Corney, D., Göker, A.: Finding newsworthy topics on Twitter. IEEE Comput. Soc. Spec. Tech. Community Soc. Netw. E-Letter **1**(3) (2013)
24. Martin, C., Göker, A.: Real-time topic detection with bursty n-grams: RGU's submission to the 2014 SNOW challenge. In: Proceedings of the SNOW 2014 Data Challenge (2014)
25. Maynard, D., Bontcheva, K., Rout, D.: Challenges in developing opinion mining tools for social media. In: Proceedings of @NLP can u tag #usergeneratedcontent?! Workshop at LREC 2012. Turkey (2012)
26. Murtagh, F.: A survey of recent advances in hierarchical clustering algorithms. Comput. J. **26**(4), 354–359 (1983)
27. Newman, N.: #ukelection2010, mainstream media and the role of the internet. Reuters Institute for the Study of Journalism working paper (2010)
28. Newman, N.: Mainstream media and the distribution of news in the age of social discovery. Reuters Institute for the Study of Journalism working paper (2011)
29. Osborne, M., Petrovic, S., McCreadie, R., Macdonald, C., Ounis, I.: Bieber no more: First story detection using Twitter and Wikipedia. In: SIGIR 2012 Workshop on Time-aware Information Access (2012)
30. Ozdikis, O., Senkul, P., Oguztuzun, H.: Semantic expansion of hashtags for enhanced event detection in Twitter. In: Proceedings of VLDB 2012 Workshop on Online Social Systems (2012)
31. Papadopoulos, S., Corney, D., Aiello, L.M.: SNOW 2014 data challenge: Assessing the performance of news topic detection methods in social media. In: Proceedings of the SNOW 2014 Data Challenge (2014)
32. Petrovic, S., Osborne, M., Lavrenko, V.: Streaming first story detection with application to Twitter. In: Proceedings of Human Language Technologies: 2010 Conference of NAACL, vol. 10 (2010)
33. Petrovic, S., Osborne, M., Lavrenko, V.: Using paraphrases for improving first story detection in news and Twitter. In: Proceedings of HTL12 Human Language Technologies, pp. 338–346 (2012)
34. Phuvipadawat, S., Murata, T.: Breaking news detection and tracking in Twitter. In: Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 3, pp. 120–123 (2010)
35. Phuvipadawat, S., Murata, T.: Detecting a multi-level content similarity from microblogs based on community structures and named entities. J. Emerg. Technol. Web Intell. **3**(1), 11–19 (2011)
36. Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Flammini, A., Menczer, F.: Detecting and tracking political abuse in social media. In: Proceedings of the International Conference on Weblogs and Social Media (ICWSM) (2011)
37. Sayyadi, H., Hurst, M., Maykov, A.: Event detection and tracking in social streams. In: Proceedings of International Conference on Weblogs and Social Media (ICWSM) (2009)
38. Schifferes, S., Newman, N., Thurman, N., Corney, D., Göker, A., Martin, C.: Identifying and verifying news through social media. Digital Journalism (2014). doi:10.1080/21670811.2014.892747

39. Shamma, D., Kennedy, L., Churchill, E.: Peaks and persistence: modeling the shape of microblog conversations. In: Proceedings of the ACM 2011 conference on Computer Supported Co-operative Work, pp. 355–358. ACM (2011)
40. Spärck, J.K.: A statistical interpretation of term specificity and its application in retrieval. J. Documentation **28**(1), 11–21 (1972)
41. Thurman, N., Walters, A.: Live blogging—digital journalism's pivotal platform? A case study of the production, consumption, and form of live blogs at Guardian.co.uk. Digital Journalism **1**(1), 82–101 (2013)
42. van Oorschot, G., van Erp, M., Dijkshoorn, C.: Automatic extraction of soccer game events from Twitter. In: Proceedings of the Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web (2012)
43. 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)

Springer
http://www.springer.com/978-3-319-18457-9

Advances in Social Media Analysis
Gaber, M.M.; Cocea, M.; Wiratunga, N.; Goker, A. (Eds.)
2015, VII, 151 p. 29 illus., Hardcover
ISBN: 978-3-319-18457-9