TOTEMSS: Topic-based, Temporal Sentiment Summarisation for Twitter

Bo Wang[1], **Maria Liakata**[1,2], **Adam Tsakalidis**[1], **Spiros Georgakopoulos Kolaitis**[1], **Symeon Papadopoulos**[3], **Lazaros Apostolidis**[3], **Arkaitz Zubiaga**[1], **Rob Procter**[1,2]
Yiannis Kompatsiaris[3]
[1] University of Warwick, UK
[2] The Alan Turing Institute, London, UK
[3] Information Technologies Institute (ITI), CERTH, Greece
{bo.wang, a.tsakalidis, m.liakata}@warwick.ac.uk

Abstract

We present a system for time-sensitive, topic-based summarisation of sentiment around target entities and topics in collections of tweets. We describe the main elements of the system and present two examples of sentiment analysis of topics related to the 2017 UK general election.

1 Introduction

In recent years social media such as Twitter have gained prominence as a rich resource for opinion mining and sentiment analysis on diverse topics. However, analysing sentiment about diverse topics and how it evolves over time in large volumes of tweets is a difficult task. In this paper, we present a system for analysing sentiment about specific topics or entities over time while providing fine-grained summaries to give insights into the underlying reasons. We illustrate its use with examples of topics discussed on Twitter during the 2017 UK general election.

Our problem formulation is related to work on prospective information needs, represented by the Microblog (Lin et al., 2015), Temporal Summarisation (Aslam et al., 2015) and Real-Time Summarisation (Lin et al., 2016) tracks at recent Text Retrieval Conferences (TRECs). However, while the aim of these tasks is to keep users up-to-date with topics of interest via push notifications or email digests, our aim is to provide an interactive user interface that shows how sentiment towards specific entities or topics develops over time. We have incorporated an automatic summarisation feature to assist users in understanding the underlying reasons. Thus, our motivation is related to the one discussed in (Meng et al., 2012), which also proposes a topic-oriented opinion summarisation framework. However, we use state-of-the-art methods enabling intuitive and interactive visualisation of sentiments in chronological order. This provides a useful tool for analysing an important event over time, such as elections, both quantitatively and qualitatively.

Here, we describe our system that aims at the aforementioned objectives. Its interactive web interface is accessible online[1]. We also present two use cases to demonstrate how the system can be used in analysing public sentiment.

2 System Design

An overview of the system is depicted in Figure 1 and comprises: 1) Data collection and sampling; 2) Sentiment classification; 3) Tweet summarisation; and 4) Data visualisation.

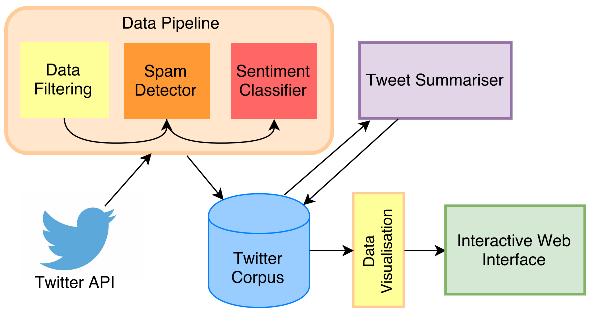

Figure 1: System overview.

2.1 Data Collection and Sampling

We collected a corpus of tweets about the 2017 UK general election through Twitter's streaming API by tracking 15 hashtags[2]. Data harvesting was performed between 26 May and 21 June 2017

[1] Live demo: http://bit.ly/2g5lBcH

[2] #ukelection2017, #ge2017, #ge17, #ukge2017, #ukgeneralelection2017, #bbcqt, #bbcdp, #marrshow, #generalelection2017, #generalelection, #electionuk, #ukelection, #electionuk2017 and #brexit

The Companion Volume of the IJCNLP 2017 Proceedings: System Demonstrations, pages 21–24,
Taipei, Taiwan, November 27 – December 1, 2017.

to capture discussions in the two weeks running up to and after the election. To identify relevant topics and entities in each tweet, we match tweets against two manually curated lists of keywords (both were created during the 2015 UK election cycle) which include 438 topic keywords relevant to nine popular election issues (e.g., immigration, NHS) and a list of 71 political party aliases (e.g. 'tories', 'lib dems'). The resulting corpus contains 3,663,090 tweets, with each tweet mentioning at least one keyword. To increase data quality and reduce noise in the corpus, we trained and applied a Twitter spam detection model using features described in (Wang et al., 2015).

2.2 Sentiment Classification

Jiang et al. (2011) showed that 40% of Twitter sentiment classification errors are caused by tweet-level approaches that disregard topics/entities. We go beyond tweet-level approaches and adopt the multi-target-specific approach proposed in (Wang et al., 2017b), which finds the syntactically connected parts of a tweet associated with each topic or entity, and extracts word embedding features from them to classify sentiment as 'negative', 'positive' or 'neutral'. This approach obtained state-of-the-art performance in both single- and multi-target benchmark data sets (Wang et al., 2017b). The whole data pipeline of Figure 1 is designed to dispatch work to many machines in parallel[3], processing many data batches simultaneously, which makes it very fast.

2.3 Tweet Summarisation

Here we aim to extract a list of representative tweets summarising the sentiment(s) expressed towards each topic/entity on each day (e.g. tweets containing positive sentiment towards 'NHS' posted on 26 June 2017).

As a prerequisite for summarisation, we group tweets containing the same sentiment towards a topic/entity on a day into a number of clusters, with each cluster assumed to represent a common theme or reason underlying the particular choice of sentiment. We adopt the two-stage hierarchical topic modelling approach proposed in (Wang et al., 2017a) and select the GSDMM+OLDA model for this task due to its effectiveness and efficiency. If there are fewer than 10 unique tweets containing the same sentiment towards a topic (or entity) on a particular day, we skip clustering and treat each of these tweets as a cluster.

To extract representative tweets summarising each cluster, we place every tweet in one common embedding space and identify 20 tweets closest to the cluster centroid (also known as metroid tweets) as summary candidates. The embedding space here is constructed using a simple but effective sentence embedding method proposed by Arora et al. (2017), which reported good performance on 22 textual similarity data sets, including a Twitter corpus. We then rank the 20 summary candidates based on weighted average tf-idf scores in the cluster; these scores can be regarded as a measure of informativeness.

We select the most informative tweet from the 20 candidates as the summary for that cluster and the final summary for the sentiment expressed towards the topic/entity is the combination of all its cluster summaries (e.g., tweets containing positive sentiment towards 'NHS' posted on 26 June 2017, comprise 8 clusters, each summarised by a single informative tweet).

2.4 Data Visualisation

For each topic/entity we calculate the following daily features: *# of tweets, # of unique users, # of tweets per sentiment type (pos, neg, neutral)* and *# of unique users per sentiment*. These features were selected on the basis of previous work on election prediction with social media (Tsakalidis et al., 2015). These are accompanied by the daily summaries of each sentiment type for a given topic/entity as described above.

In addition to showing the raw values of the above features, we also normalised sentiment features (*# of tweets per sentiment, # of unique users per sentiment*) to reflect the percentage of sentiment of a particular type towards a topic/entity on a particular day. To allow time series comparisons across different topics/entities we normalised the *# of tweets* and *# of unique users* of all topics/entities across all days in the range [0, 1]. Finally, to account for differences in popularity, we calculated the average (per-topic and across all days) *# of tweets* and *# of unique users*.

The web interface is implemented on Web standards (HTML5/CSS3). The timeline graphs are built using the NVD3[4] library (reusable charts for `d3.js`), while the auto-complete function-

[3]We ran it on a server with 40 CPU cores and 64 GB of RAM.

[4]http://nvd3.org/

ality is based on the 'Ajax AutoComplete for jQuery' library[5]. In addition, jQuery from Google Hosted Libraries[6] and D3.js from Cloudfare Hosted Libraries[7] are used for DOM manipulation (add/remove elements, click events, etc.) and accessing data (from tsv files) respectively.

3 Example Use Cases

We use two use cases to demonstrate how our system can help analyse public sentiment on Twitter.

3.1 Use Case #1 – Party Sentiment

Recent election campaigns suggest that the Twittersphere tends to contain more negative sentiment during the election period. Hence, in the first case study, we compare negative sentiment trends on Twitter for the two major UK political parties, 'Conservative' and 'Labour', before and after the 2017 UK general election. As described in section 2.4, the negative sentiment reflects the percentage of negative sentiment for each party on each day over all sentiment bearing tweets.

Figure 2 reveals consistently more negative sentiment towards 'Conservative' than 'Labour', especially for the week before election day (8 June).

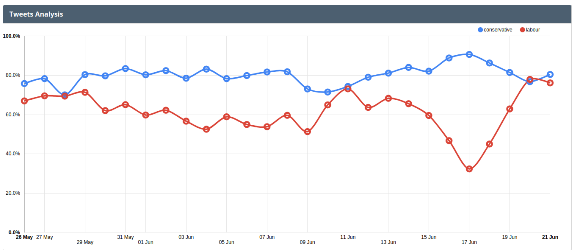

Figure 2: Negative sentiment trends for 'Labour' (red) and 'Conservative' (blue).

3.2 Use Case #2 – Grenfell Tower Fire

To provide deeper insight into the advantages of our opinion summarisation system, we present a case study on how public sentiment towards the topic 'housing' developed before and after the Grenfell Tower Fire disaster[8]. Figure 3 shows the percentage of users expressing negative sentiment towards 'housing' as well as the governing party 'conservative' over the period covering the incident. Our web interface allows users to click on each circle shown on the graph to display tweet summaries for that topic on that particular day.

We can see the number of users expressing negative sentiment for the topic 'housing' fluctuated throughout the election period while it remained fairly constant for 'Conservative'. Negative sentiment peaked in both cases on June 16th.

Table 1 presents a negative sentiment summary for each day between June 12 and 15, and all three negative opinion summary tweets on the peak day of June 16 showing each summary tweet represents a different aspect of the topic. Along with the graph shown in Figure 3, this summary offers a tight integration of topic, sentiment and insight into reasons behind the sentiment. Before the fire, negative sentiment towards 'housing' was austerity related; after the fire, the incident dominated the 'housing' discussion on Twitter. A large portion of users blame the Conservative government for the decline of social housing and ultimately the Grenfell Tower fire. Finally, on June 16 each of the negative opinion summaries represents one theme related to this topic, namely 'the decline of social housing', 'immigration and housing' and 'the votes on housing safety'.

4 Conclusion

We presented a monitoring system for topic-entity sentiment on Twitter that summarises public opinion around the sentiment towards each entity. The system deployment for the 2017 UK election, provides an interactive visualisation for comparing sentiment trends and display opinion summaries on the graph. In the future, we plan to improve our system to produce more concise summaries and allow near real-time processing of new events.

References

Sanjeev Arora, Yingyu Liang, and Tengyu Ma. 2017. A simple but tough-to-beat baseline for sentence embeddings. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*.

Javed Aslam, Fernando Diaz, Matthew Ekstrand-Abueg, Richard McCreadie, Virgil Pavlu, and Tetsuya Sakai. 2015. Trec 2015 temporal summarization track overview. In *Proceedings of the 24th Text REtrieval Conference, TREC*.

Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. 2011. Target-dependent twitter senti-

[5] https://www.devbridge.com/sourcery/components/jquery-autocomplete/
[6] https://developers.google.com/speed/libraries/
[7] https://cdnjs.com/
[8] https://en.wikipedia.org/wiki/Grenfell_Tower_fire

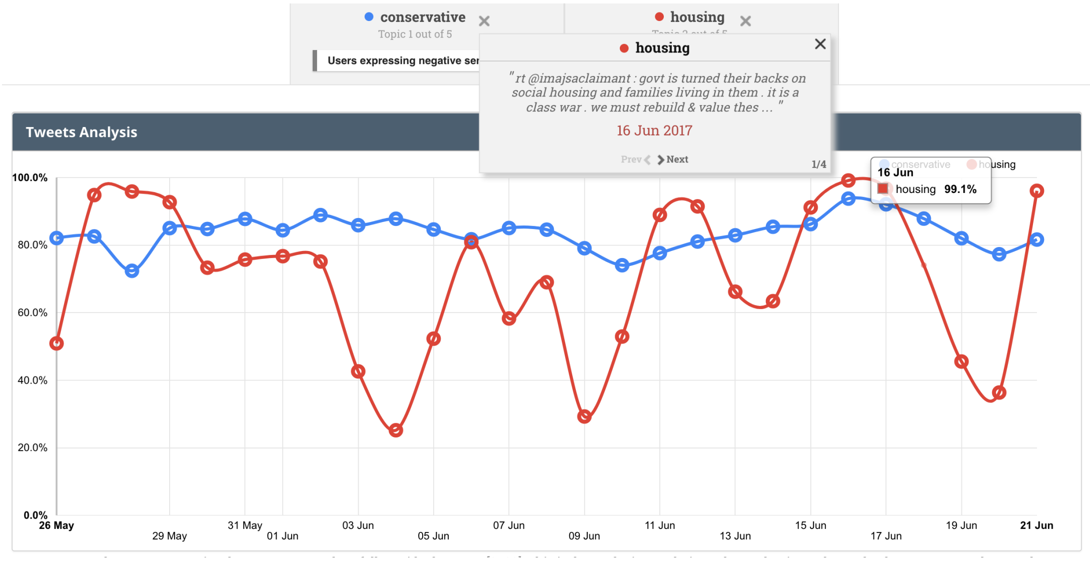

Figure 3: Negative sentiment trends for 'housing' (red) and 'conservative' (blue), with a summary tweet displayed for the former.

Topic entity	Opinion Summaries	Date
housing	rt @user1 : the audacity to even refer to tackling a " housing crisis " after being in government for 7 years . https://t.co/lifwybhryp	12 June 2017
housing	austerity is still here , bedroom tax , foodbanks , pip , housing cap , universal credit taper , welfare freeze , esa cuts , inflation is up . #ge17	13 June 2017
housing	@bbcnews @skynews @itvnews tories cuts in society kill just look at social housing #grenfelltower sold to cheapest bidding #ge17 #bbcqt	14 June 2017
housing	tory capitalism cutting kills social housing on the cheap #grenfelltower cuts in fire ambulance police nhs services #victorialive #ge17	15 June 2017
housing	rt @user2 : govt is turned their backs on social housing and families living in them . it is a class war . we must rebuild & value thes ...	16 June 2017
housing	rt @user3 : laura perrins again blaming the death toll of #grenfelltower on immigration - putting pressure on housing . laura bt ...	16 June 2017
housing	rt @user4 : it is a shame the ministers hearts did not go out to the people in grenfell tower when they were voting on housing safety #bbcqt	16 June 2017

Table 1: Negative opinion summary for 'housing' before and after the Grenfell Tower fire

ment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*.

Jimmy Lin, Miles Efron, Yulu Wang, Garrick Sherman, and Ellen Voorhees. 2015. Overview of the trec-2015 microblog track. In *Proceedings of the 24th Text REtrieval Conference, TREC*.

Jimmy Lin, Adam Roegiest, Luchen Tan, Richard McCreadie, Ellen Voorhees, and Fernando Diaz. 2016. Overview of the trec 2016 real-time summarization track. In *Proceedings of the 25th Text REtrieval Conference, TREC*, volume 16.

Xinfan Meng, Furu Wei, Xiaohua Liu, Ming Zhou, Sujian Li, and Houfeng Wang. 2012. Entity-centric topic-oriented opinion summarization in twitter. In *Proceedings of the 18th ACM SIGKDD*.

Adam Tsakalidis, Symeon Papadopoulos, Alexandra I Cristea, and Yiannis Kompatsiaris. 2015. Predicting elections for multiple countries using twitter and polls. *IEEE Intelligent Systems*, 30(2):10–17.

Bo Wang, Maria Liakata, Arkaitz Zubiaga, and Rob Procter. 2017a. A hierarchical topic modelling approach for tweet clustering. In *Proceedings of the 9th International Conference on Social Informatics (SocInfo)*.

Bo Wang, Maria Liakata, Arkaitz Zubiaga, and Rob Procter. 2017b. Tdparse-multi-target-specific sentiment recognition on twitter. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*.

Bo Wang, Arkaitz Zubiaga, Maria Liakata, and Rob Procter. 2015. Making the most of tweet-inherent features for social spam detection on twitter. In *5th Workshop on Making Sense of Microposts (#Microposts2015) WWW*, volume 1395, pages 10–16.