NIA: System for News Impact Analytics

Mikalai Tsytsarau
University of Trento
tsytsarau@disi.unitn.eu

Themis Palpanas
Paris Descartes University
themis@mi.parisdescartes.fr

ABSTRACT

The analysis of news impact on people is relevant to a variety of applications, ranging from monitoring product and companies reputations, to stock market prediction. Therefore, it is important to understand the underlying mechanisms which affect the propagation of news and drive the evolution of sentiments in one way or another. In this demonstration paper, we describe `NIA`, a system that identifies and describes news events that caused changes of sentiments. `NIA` is based on a novel framework for a complex news event modeling, which is capable of detecting time and importance characteristics of events by only observing a time series of news articles publications, and then correlating this data with a time series of sentiment shifts. The operation of our system is summarized as follows. First, we apply a deconvolution to recover the time, longitude, importance and impact of news events. Second, we compute a sentiment time series, e.g., by monitoring sentiments for positive or negative bursts, and coherently analyze sentiment and news time series, automatically determining their time lag. Third, we evaluate the corresponding news articles for a time interval of interest and extract the essence of what happened. Finally, we present the selected news time series to the user, as well as several more correlated stories, which could have affected sentiments as well, proposing to interactively explore their connections.

1. INTRODUCTION

Today, sentiment analysis has become a platform that provides valuable information on people's opinions regarding different topics, and is widely used by businesses and social study institutions [6]. By aggregating sentiments, expressed in multiple texts, and assessing the result with statistical measurements, we can capture certain changes, or shifts, in global sentiment, which cannot be attributed to random variation [5]. Recent studies indicate that the observed sentiment changes can be the result of people reacting differently to external events [4, 7], opening this problem for the investigation.

In this demo, we aim at determining the impact of news events on sentiment changes. However, most of the news events are announced as atomic pieces of information and their importance is not readily intelligible from the text alone. To determine the importance and impact of news to people, it is crucial to consider the relevant publication dynamics of the whole crowd, rather than only from news agencies or news media. Whats more, it is important to analyze all types of sentiment shifts, which could be connected with news events. The problem is that relevant sentiment shifts can be particularly small and can occur before, during or after the event - all with varying delays, depending on the event type and publication pace of the media. This necessitates the sophisticated news and sentiment extraction, aggregation and tracking methods, as well as proper correlation measures between news and sentiments.

Such problems require processing significant amounts of data to produce a desired output, from sentiment extraction to event processing. However, the most challenging part of our problem is finding relevant pairs of news events and sentiment shifts, because there is usually no one-to-one correspondence between the event and sentiment shift types, and there can also exist multiple correlated topics, which contribute to sentiment deviations. At this step, the interaction with the user in order to pick up such cases can be very beneficial, since the system can quickly filter through the correlated topics, but only human can understand the semantic connection (and causality) between events and sentiments.

This demonstration features `NIA` [7] - a system for news and sentiment analytics, which monitors important news events, evaluates their dynamics, and captures the correlated sentiment changes. Our system aims to predict which event types are likely to cause the sentiment to change by analyzing news importance and dynamics and letting the user to explore the connections between time series of sentiment shifts and news events for correlated topics.

The `NIA` system relies on principled techniques and approaches to news and sentiment aggregation and analysis: (a) we employ automated parameter optimization for processing time series, detecting news events and measuring their characteristics; (b) sentiment noise and irregularity are reduced by regression smoothing, taking into account the diversity and significance of sentiments.

Motivating Scenarios and Examples:

Example 1: We want to detect changes in the opinion on a particular topic, when such changes are caused by news events. For instance, imagine the situation demonstrated in Figure 1, in which the sentiment expressed in Twitter for the Large Hadron Collider (LHC) has dropped from positive to negative just after the first experiments begun. In our example, we see that people started to talk negatively in the aftermath of the first experiments (marked *"collision"*), while the news about the record beam energy (marked *"record energy"*) pushed sentiments back to neutral. To understand the difference between these two events we need to navigate to a correlated news trend and analyze the volume of news around these sentiment changes. However, proper news event detection and processing require special methods, as shown in our next example.

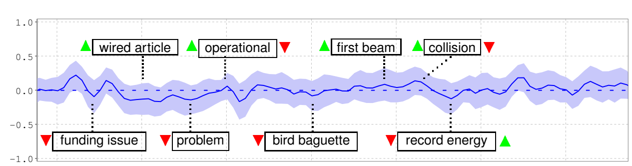

Figure 1: Sentiment shifts for the topic "LHC" from Twitter.

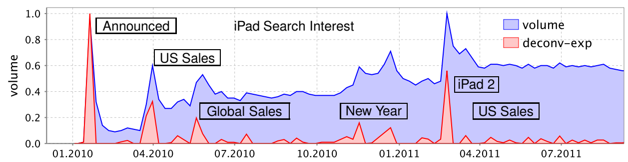

Figure 2: Events identified by deconvolution for "iPad".

Example 2: Consider a *search interest* time series extracted from Google for the topic "iPad", shown in Figure 2, blue. It features a growing number of search queries overlaid with a series of overlapping bursts of user interest, making it very hard to detect news events. For instance, the relative difference in interest between *"iPad 2"* and the following *"US Sales"* events is obstructed by the trend, which makes their volumes appear similar. The output time series of events (Figure 2, red), processed using our method, demonstrates a more vivid event separation, making them easily detectable. Moreover, it appears without the global trend, revealing true event importance and dynamics.

2. NEWS IMPACT ANALYTICS

Our system for news impact analytics, `NIA` [7], addresses the problems of detecting interesting changes of aggregated sentiment and connecting them to relevant news events which could have caused these situations. In this section we briefly introduce the main capabilities and design principles of `NIA`, proceeding with the description of its main features and the demonstration scenario in Section 3.

2.1 System Overview

Figure 3 outlines the composition of `NIA`. It consists of *Sentiment Analysis* and *News Event Analysis* layers, which analyze aggregated sentiment data and news volume as described below. The sentiment analysis layer takes care of aggregating sentiments for a topic and detecting interesting changes, which can be contradictions, outbursts of sentiments' volume or other changes in sentiment happening over time. The news event analysis layer works with time series of publication volume (which can be news, blog posts, tweets) to detect various events that could have caused the observed shifts in sentiment. Events are detected with the help of deconvolution, through observing outbursts in news volume, and are annotated automatically by summarizing relevant news articles.

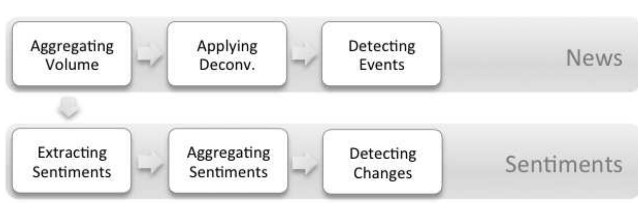

Figure 3: Compositional diagram of the system.

2.2 Sentiment Analysis

We determine topic T and sentiment S for each text and assign a continuous sentiment value S in the range [-1;1] that indicates the polarity of the opinion expressed regarding the topic. For the sentiment assignment step, we use the SentiStrength [4] tool, which recognizes opinion expressions, emoticons and works especially well for short texts, like tweets.

For analyzing news impact, we are interested in sentiment measures that are sensitive to particular kinds of sentiment changes, usually correlated with events. However, not many studies propose suitable measures for opinion shifts, which can be analyzed coherently with the news time series in order to extract correlations. The particular methods which can be adopted to our problem are *sentiment volume* [4] and *contradiction level* [8], discussed below.

Sentiment Volume is defined as the amount or the sum of sentiments of a particular polarity, expressed within a specified time interval [4]. It captures bursts of particular opinions, e.g., *positive*:

$$s(t) = \sum_{i=1}^{n} S_i^+(t), \quad \text{or} \quad s(t) = |S_i^+|(t)$$

Contradiction Level is another suitable measure for sentiment shifts, that can detect both changes of sentiment polarity as well as temporary shifts of sentiments [8]. The intuition behind this measure is that when the aggregated value for sentiments μ_S is close to zero, while the sentiment diversity (variance) σ_S^2 is high, then the contradiction should be high. Combining μ_S and σ_S^2 in a single formula, we propose the following measure for contradictions:

$$s(t) = \frac{\vartheta \cdot \sigma_S^2}{\vartheta + (\mu_S)^2} W(n),$$

where n is the number of sentiments, $\vartheta \neq 0$ is the normalizing constant, and W is a weight function that takes into account the significance of sentiment statistics involved in the calculation [8].

2.3 News Event Analysis

We consider that sentiment changes can be preceded with or followed by news events. A time lag between the two sequences can be determined by maximizing their cross-correlation coefficient. It can then be used to navigate to the relevant news event, given a time interval of sentiment shift, annotating it with the keyword description and importance dynamics.

Extracting News Time Series. An example time series of news volume is shown in Figure 2. It consists of a series of bursts of varying height and length, which can even be overlapping. Constructing the news volume time series $n(t)$ for a specific topic involves the analysis of documents in the collection $\mathscr{D}$ and estimation of topic's popularity (frequency) among them. For example, we can count a number of documents D_i which have occurrences of the topics' keywords T, or sum their TF-IDF scores:

$$n(t) = |\mathscr{D}^T|_t = \{D_i \in \mathscr{D} \mid T \in D_i\}; \quad n(t) = \sum_{D_i \in \mathscr{D}^T} \mathit{TF\text{-}IDF}(T, D_i)|_t$$

Detecting Impacting Events. As we already noted, not every kind of publications outbursts is caused by external news, and not every kind of news dynamics has an impact on sentiment, so we want to distinguish them at a fine level of detail, for instance, distinguishing between *linear*, *hyperbolic* or *exponential* response types, either symmetric or asymmetric, depending whether events are anticipated or not. Our system represents news publication volume as the result of the interplay between the original news' importance $e(t)$ and media response $mrf(t)$, in a process known as convolution. In order to recover $e(t)$, we perform a deconvolution of news volume time series, using Fourier transformation, as described in [7]:

$$e(t) = \mathscr{F}^{-1}\{e(\omega)\} = \mathscr{F}^{-1}\{n(\omega)/mrf(\omega)\}$$

Unlike other models [2, 1, 3], describing publication dynamics by complex equations, deconvolution uncovers *succinct* and *meaningful* event parameters in the form of $e(t)$, such as: event's interest *buildup and decay*, its *longitude* and *maximum importance* level.

Extracting Event Annotations. To automatically annotate news event, we compare TF-IDF scores of the news documents within a current time interval to the same scores over the entire collection of news, and extract top k terms, which became more popular in the event time interval $\mathscr{D}_e^T$:

$$T_{event} = \{T_j \mid max_k(\textit{TF-IDF}(T_j, \mathscr{D}_e^T) - \textit{TF-IDF}(T_j, \mathscr{D}^T))\}$$

Correlating News and Sentiments.

We observe that sentiment and news time series require special correlation methods, that are different to conventional Pearson cross-correlation coefficient, which measures the linear dependency between variables. Such time series do not have a definite average level, around which the movement is happening. Instead, their values are outbursting from the minimum level at particular points in time. Therefore, we apply binary similarity measures, for example *cosine similarity* or *Jaccard coefficient*, measuring the intersection between sentiment and event bursts. In addition to counting the number of overlapping bursts, we can apply their weighting, for example based on magnitude.

3. DEMONSTRATION SCENARIO

Our system is capable to detect sentiment shifts in multiple time series and correlate them with news events in real time. In this demonstration, we intend to show the main features of our system on the real dataset from Twitter, by applying `NIA` on the stored data flow and giving users a possibility to visualize and explore news events, along with their sentiment changes, automatically extracted in real time. The important feature of our system is that it assigns sentiment changes to events on the same topic automatically based on their correlation, and also allows user to explore and suggest events from other related topics.

Demonstration Dataset. For our demonstration dataset, we selected 30 trending topics from Twitter, which featured the most prominent events for the period of half a year, from June 2009 till December 2009. The dataset contains approximately 7 million tweets in total and over 400 peaks during the events. We use 1-day aggregation for the time series of tweets volume and sentiments.

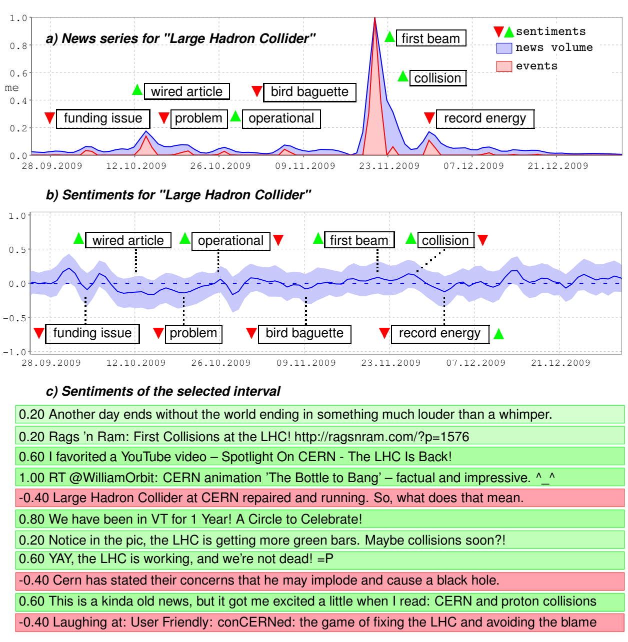

Figure 4: NIA demo workflow for the topic "LHC" from Twitter.

Demonstration Workflow. We intend to demonstrate an interactive application, shown in Figure 4, which allows users exploring news time series for each of the topics in our dataset, visualize the corresponding sentiments, drill down to the actual positive and negative posts, and see which other relevant news events could have affected sentiments, based on correlation analysis. Users can interact with the system by selecting and zooming time series, and also by adjusting various parameters, such as aggregation granularity, smoothing level and correlation thresholds, in real time.

Our demo starts by displaying to users a graph with the news volume, as seen in Figure 4(a). In this graph, `NIA` automatically extracts and annotates the relevant news events. Moreover, it marks the related sentiment shifts near text event annotations. The user can also visualize the entire time series of average sentiment, contradiction level, positive or negative sentiment volume, which in this case also become annotated with sentiment shifts and event labels, shown in Figure 4(b). In cases, when events cause transitions of sentiments (from positive to negative or vice verse), event annotations are marked with the two corresponding arrows, as seen in events marked as *"collision"* and *"record energy"*. Finally, by clicking on the interesting time interval, users are able to see a time series of posts, marked with positive (green) and negative (red) sentiment labels, as shown in Figure 4(c) for the event *"first beam"*.

4. CONCLUSIONS

Our system allows correlated analysis of sentiments and news, and raises new data analysis opportunities, useful for sociology and marketing researchers. Our evaluation reveals the existence of different parameters for various events, even for the same topic, all having different impacts on sentiments, suggesting that it is possible to predict sentiment changes. To achieve this, we need to take into account the type of response dynamics in addition to the event's importance level, creating a more elaborate causality model. This task requires building a database of event and sentiment shift profiles, and exploration of events on related topics, in addition to events on the same topic, leading to the necessity for employing a sophisticated and interactive analytics platform, which helps users in their search for event causality. The purpose of our demo system is to facilitate the development of such a platform and explore possible ways of interactive news and sentiment analysis.

5. REFERENCES

[1] S. Asur, B. A. Huberman, G. Szabó, and C. Wang. Trends in social media: Persistence and decay. In *ICWSM*, 2011.

[2] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *KDD*, 2009.

[3] Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos. Rise and fall patterns of information diffusion: Model and implications. In *KDD*, pages 6–14, 2012.

[4] M. Thelwall, K. Buckley, and G. Paltoglou. Sentiment in twitter events. *JASIST*, 62(2):406–418, 2011.

[5] M. Tsytsarau, S. Amer-Yahia, and T. Palpanas. Efficient sentiment correlation for large-scale demographics. In *SIGMOD*, pages 253–264, 2013.

[6] M. Tsytsarau and T. Palpanas. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery, Special Issue on 10 Years of Mining the Web*, pages 1–37, 2011.

[7] M. Tsytsarau, T. Palpanas, and M. Castellanos. Dynamics of news events and social media reaction. In *KDD*, 2014.

[8] M. Tsytsarau, T. Palpanas, and K. Denecke. Scalable detection of sentiment-based contradictions. In *DiversiWeb*, 2011.