

Large-scale Comparative Sentiment Analysis of News Articles

F. Wanner, C. Rohrdantz,
F. Mansmann, A. Stoffel, D. Oelke,
M. Kristajic, D. A. Keim*
University of Konstanz, Germany

D. Luo, J. Yang[†]
University of North Carolina Charlotte, USA

M. Atkinson[‡]
Joint Research Centre of
the European Commission
in Ispra, Italy

ABSTRACT

Online media offers great possibilities to retrieve more news items than ever. In contrast to these technical developments, human capabilities to read all these news items have not increased likewise. To bridge this gap, this poster presents a visual analytics tool for conducting semi-automatic sentiment analysis of large news feeds. The tool retrieves and analyzes the news of two categories (Terrorist Attack and Natural Disasters) and news which belong to both categories of the Europe Media Monitor (EMM) with respect to positive and negative opinion words. While this happens automatically, the more demanding news analysis of finding trends, spotting peculiarities and putting events into context is left to the human expert.

Index Terms: H.5.2 [Information Interfaces and Presentation]: Miscellaneous—

1 INTRODUCTION

Analyzing news stories and user generated content is of huge importance for many people and organizations, such as politicians who want to find out their public reputation or reviews about products can considerably influence sales volumes.

Since there is a huge amount of news every day, our goal is to offer a semi-automatic approach by taking news data from the Europe Media Monitor [1], conducting sentiment analysis on the news to assess how positive or negative a particular news postings is, and then to present the information in a visual analysis tool. While our approach is not suitable to completely replace a thoroughly conducted opinion poll due to the lack of accuracy, it has also some unique advantages, namely low costs and the possibility to continuously monitor a particular subject in real-time.

In this paper, we demonstrate a way of using text analysis methods in combination with a novel visual representation. On the one hand, this system automatically evaluates the emotional content of a news item. On the other hand, the visual interface empowers the human expert to draw meaningful conclusions, to selectively read a few news postings with strong emotional content, to discover trends, and to gain an overview of the development of chosen topic in the media.

2 VISUAL SENTIMENT ANALYSIS

2.1 Data Processing

The data we used was gathered from the news of the Europe Media Monitor (EMM). The EMM news feed informs about news published on websites in different languages from countries all over the world and automatically annotates the news with meaningful

*e-mail: {franz.wanner, christian.rohrdantz, andreas.stoffel, daniela.oelke, florian.mansmann, milos.kristajic, daniel.keim}@uni-konstanz.de

[†]e-mail: {dluo2, Jing.Yang}@uncc.edu

[‡]e-mail: martin.atkinson@jrc.ec.europa.eu

4 weeks

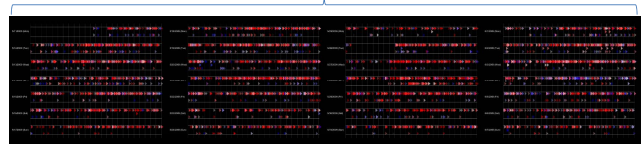


Figure 1: The visualization showing a time period of four weeks.

categories. For our analysis we considered only news in English and focused on news items from the categories “Natural Disasters” and “Terrorist Attack”, which play a crucial role in the civil security field. In one month we collected about 1000 news items reporting about natural disasters and about 6000 news items reporting about terrorist attacks. For every news item we process date, source, title and description and moreover enrich it with a sentiment score. For this purpose we make use of a freely available list of words that evoke positive or negative associations [2]. We count the number of positive and negative words in a news item and finally the absolute relation of positive against negative words provides our sentiment score for this item.

2.2 Data Visualization

The visualization on the one hand aims to give a meaningful representation of the data and on the other hand is intended to be an appropriate starting point for the interactive exploration and discovery of interesting patterns. Figure 1 shows a screenshot of the visualization. Each of the four major blocks represents the news of one week. A week is split up into single days along the vertical axis. Each group of two lines represents one day and each colored object depicts one news item as you can see in Figure 2 and Figure 4. The category of a news item determines if it is placed at the upper line or lower line of a day or in between. In addition, a news item’s sentiment score is encoded by its color hue and saturation. The following passages describe each of those aspects in detail.

2.2.1 Placement

Every news item is represented by a triangle in a 2D plane. The position of the object within the plane depends on the date the news was published and on its category (see figure 3). Thereby, the day it was published accounts for the pair of lines it will be placed in between (as each line pair represents one day) and the time of day determines its horizontal position within the line. The exact vertical position depends on the category a news item belongs to: The upper line of a day contains all news items belonging to the first category - here terrorist attack - and the lower line contains the news of the second category - here natural disasters. News items that belong to both categories are placed in between both lines.

2.2.2 Coloring & Shape

News items with a positive sentiment score get a blue color hue and negative news items appear in red. Thereby, the color saturation

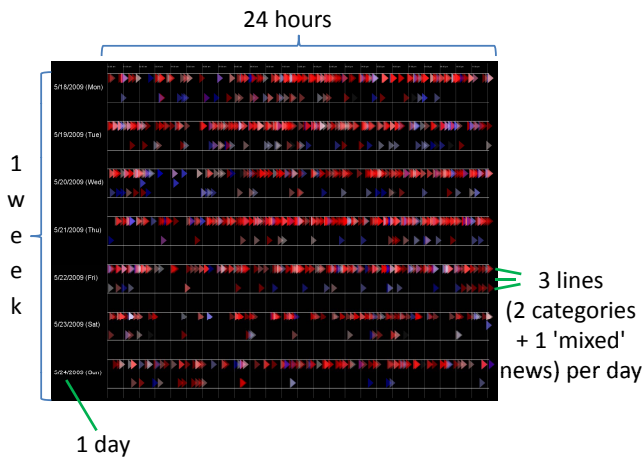


Figure 2: The arrangement of news within one week.

indicates the strength of the previously determined sentiment polarity. That means news items with higher sentiment scores get more saturated colors and therefore become salient.

At the same time, we reduced the opacity of the news items which has two desired effects. The semi-transparency of the items reduces the problem of high overlap among the news because single items in common cases can still be distinguished. Additionally, a cluster of news items that simultaneously report about the same event with the same polarity thus will stand out. In the overlapping areas news items will mutually amplify their opacity resulting in a stronger color.

Every news item has a triangle shape. This has the implication that the vertical space a news item covers is at its maximum when it is published and then is slightly vanishing. This is congruent with the influence news exert on analysts: They are important when they appear and then become less interesting as new events arise. Moreover, the triangle shape allows an analyst to distinguish single news items at the upper and lower apex even if they are strongly overlapping. In contrast, in the middle of each triangle a larger overlap to neighboring items is provided in order to enable the effect of visual aggregation through mutual opacity amplification.

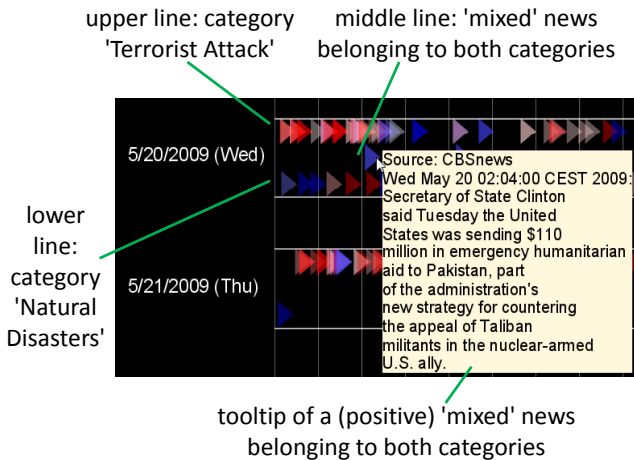


Figure 3: Positioning of the news in the visualization depends on the date and category.

2.3 Interactive Visual Analytics

The visualization is designed for an interactive data exploration. There are several possibilities to interact with the tool. Continuous zooming allows to analyze certain parts at a greater level of detail. From a certain zoom level on, the horizontal scale of the news triangles is reduced while the background scale is still enlarged. This has the desired effect that the triangles are not simply getting constantly larger but become separated when a further enlargement would not reveal additional insights. Thus, there always is a zoom level where each single news item will be displayed without overlap in order to allow a more in-depth analysis for a certain time interval, as can be seen in figure 4.

Of course, panning is also enabled in the different zoom stages to facilitate the exploration of neighboring regions. Additional details are provided on demand - when the mouse is dragged over a triangle, a tooltip appears containing date, time, original news source, and description of the item.

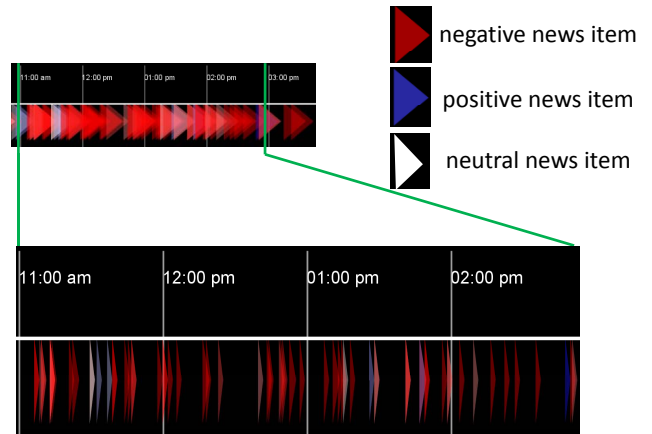


Figure 4: Non-overlapping zoom for a in-depth analysis for a certain time interval.

3 CONCLUSIONS

In this paper we presented a visual tool for large-scale comparative sentiment analysis of news articles. Thereby, we see two research contributions: 1) We retrieved a live-stream of news elements in XML from the Europe Media Monitor and processed it by assigning sentiment scores using a keyword-based heuristic, which is capable of understanding negation. 2) The zoomable user interface visualizes news articles as arrow heads, which are colored according to positive or negative sentiment scores. Aligning several of news categories above each other enables a) a rough quantitative assessment of the number of articles in each category, b) an overview of the publishing time of the articles, and c) a comparison of the categories' sentiments.

We believe that this approach empowers media analysts to effectively filter out emotional news stories, which can be used in a large number of research and business cases.

REFERENCES

- [1] M. Atkinson and E. Van der Goot. Near Real Time Information Mining in Multilingual News. In *Proceedings of the 18th International World Wide Web Conference (WWW'2009)*, pages 1153–1154, 2009.
- [2] V. Buvac. Internet General Inquirer, 2008. <http://www.webuse.umd.edu:9090/> as retrieved on Nov. 14, 2008.