



HAL
open science

Visualizing Time-varying Twitter Data with SentimentClock

Florence Ying Wang, Arnaud Sallaberry, Karsten Klein, Masahiro Takatsuka

► **To cite this version:**

Florence Ying Wang, Arnaud Sallaberry, Karsten Klein, Masahiro Takatsuka. Visualizing Time-varying Twitter Data with SentimentClock. InfoVis: Information Visualization, 2014, Paris, France. IEEE Information Visualization Conference (InfoVis 2014), Poster Abstracts of IEEE VisWeek 2014, 2014. lirmm-01275390

HAL Id: lirmm-01275390

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-01275390>

Submitted on 17 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Visualizing Time-varying Twitter Data with *SentimentClock*

Florence Ying Wang*
LIRMM, University Montpellier 2
The University of Sydney

Arnaud Sallaberry†
LIRMM, University Montpellier 2

Karsten Klein‡
The University of Sydney

Masahiro Takatsuka§
The University of Sydney

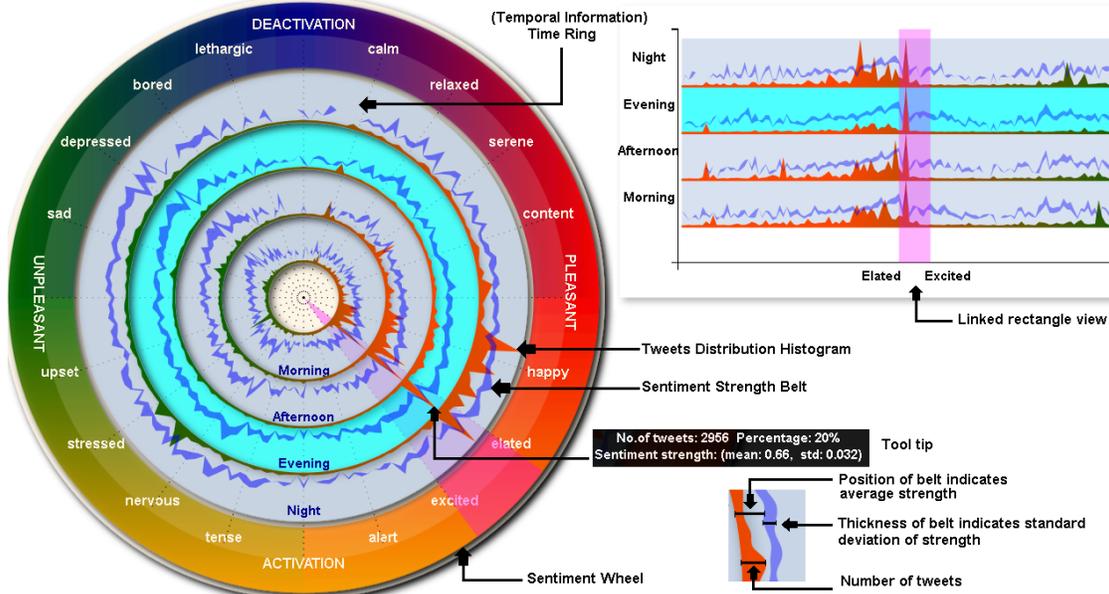


Figure 1: *SentimentClock* of the tweets collected on 2013 Australian election day (7-Sep-2013).

ABSTRACT

Microblogs such as tweets contain useful information for sentiment analysis. In this work, we introduce *SentimentClock* for visualizing the sentiment of time-varying Twitter data on 2D affective space. To illustrate our visualization design, two case studies are conducted. They demonstrate the effectiveness of *SentimentClock* in visualizing the temporal sentiment variations of tweets and comparing the sentiment of tweets on different topics. A demo video of our system is available at: <http://youtu.be/JvQFAFW-VbI>

Index Terms: H5.2 Information System and Interfaces.

1 INTRODUCTION

The growth of online microblogging services, particularly Twitter, have enabled people to express and spread opinions by posting short text messages such as “tweets”. Therefore, the collection of tweets contains an enormous amount of information that are related to people’s opinions and sentiments on various topics. Given such an advantage, many research studies use tweets as a basis for sentiment analysis. Usually, the analysis aims to gauge people’s opinions (e.g. positive or negative) or complex emotional states (e.g. happy, sad,

stressed) conveyed by their tweets using text analysis methods.

To visualize the sentiments of tweets, many previous works, e.g. [3], abstracted “sentiment” as a unidimensional variable, e.g. from negative to positive. However, recent psychological studies underline the more complex nature of sentiments. For instance, “sentiment” in the circumplex model [2] are considered to form a multidimensional polar space, i.e. the affective space, having two principal axes: valence (i.e. from unpleasant to pleasant) and arousal (i.e. from deactivation to activation). Therefore, the unidimensional abstraction of sentiment cannot truly represent the distances between different emotional states. To overcome this, Ramaswamy [4] combined scatter plot, heat map and tag cloud with the affective model proposed by [2] for visualizing the two dimensional sentiment of tweets using multiple views. With his visualization, the representation of “sentiment” has been improved, however, the temporal information of tweets are not associated with the visualization of 2D sentiment. Therefore, temporal variation of 2D sentiments of tweets (modeling complex emotions expressed by tweets) cannot be visualized. In this work, we aim to address this problem by proposing a visualization design called *SentimentClock*.

2 TASKS

Many interesting tasks related to sentiment analysis of microblogs require the consideration of the temporal dimension. We have compiled a list of these tasks and use them as the guidelines for designing and evaluating our visualization:

- [T1] Visualize and compare temporal variations of sentiments for collected tweets.
- [T2] Compare sentiments variations of tweets on different topics.
- [T3] Visualize the distribution of tweets on 2D affective space.

*e-mail: florence8627@vislab.net

†e-mail: arnaud.sallaberry@lirmm.fr

‡e-mail: karsten.klein@sydney.edu.au

§e-mail: masa.takatsuka@sydney.edu.au

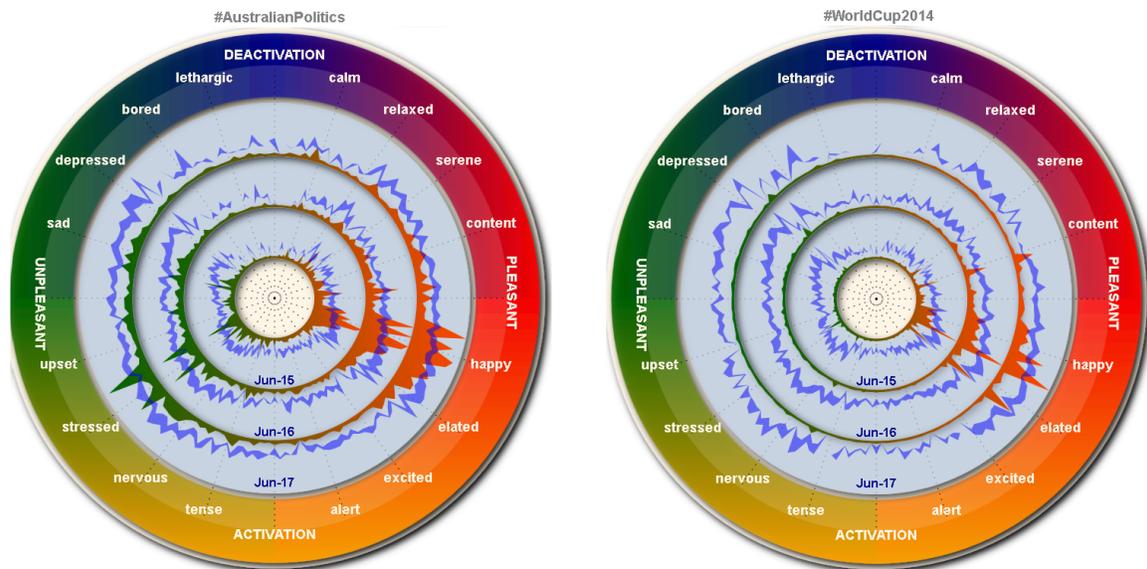


Figure 2: Comparing the sentiment variations of tweets on two topics: “Australian Politics” (left) and “World Cup 2014” (right).

[T4] Visualize both dimensions of sentiments (i.e. valence, arousal) of tweets and their semantic meanings (e.g. elated, stressed).

3 DATA ABSTRACTION

We adopt the ANEW dictionary [1] to estimate sentiment of tweets. It contains affective ratings for 1034 English words obtained from empirical studies. For each word, the ratings consist mean and standard deviation for both valence and arousal. To calculate the 2D sentiment (i.e. valence and arousal) of tweets, two types of signaling words are searched: (1) affective words recorded in ANEW; (2) negation words, e.g. “should not, don’t”. We use the method suggested by [4] to statistically weight the ratings of multiple affective signaling words in a tweet. Meanwhile, if a negation word exists, we estimate the sentiment to be opposite to what are indicated by affective words. Eventually, for each tweet, polar coordinates consisting of the angle specifying a sentiment (Russell’s model [2]), and the radius indicating the strength of the sentiment are obtained.

4 VISUAL MAPPING

The features of our visualization design are shown in Figure 1.

(1) For [T1, T2], *Time rings* are used for adding the temporal dimension. *Linked rectangle view* and *Tool tips* showing statistics are used to allow better comparison.

(2) For [T3], *Tweets distribution histogram* is calculated based on the angle between valence and arousal. Color of the histogram shows positivity (red) or negativity (green) of the sentiment, which is mainly associated with valence [2, 4]. *Sentiment strength belt* shows the strength of sentiment (i.e. radius under polar coordinate system). Position of the belt shows the average strength. Thickness of the belt indicates standard deviation of such strength.

(3) For [T4], *Radial layout* is used to follow Russell’s affective model [2]; *Sentiment wheel* is labeled with affective words and maps two pairs of complimentary colors to valence and arousal, i.e. green to red shows pleasant level (valence), blue to yellow shows activation level (arousal).

5 CASE STUDIES

To illustrate our visualization, we collected tweets via Twitter streaming API for two case studies.

Case study 1 Our first case study is based on tweets collected during 2013 Australian election period. 36016 related tweets posted

on the election day (0:00 to 24:00, 7th Sep, 2013, GMT+10) are selected to visualize the temporal variation of sentiments. As shown by Figure 1, sentiments of tweets in the morning and afternoon (0:00 to 18:00) follow similar distribution ([T1, T3]). In the evening (18:00 to 22:00), which is the vote counting and result releasing period (highlighted in blue), tweets are found to have both high arousal and high valence, primarily falling into the elated and excited range (highlighted in purple) with high strength low standard deviation shown by the strength belt and the tool tip ([T1, T4]). At night (22:00 to 24:00), which is just after result announcement, sentiments split into two parts (with more being happy and elated) but both having high arousal ([T3, T4]).

Case study 2 Our second case study is based on tweets collected from 15th Jun to 17th Jun, 2014 tracking two topics: “Australian Politics” and “World Cup 2014”. Around 71200 tweets are found to contain affective signaling words for sentiment visualization. As shown by Figure 2, tweets on the topic “Australian Politics” are more spread out along the sentiment wheel and express more negative sentiments, e.g. upset and stressed, than the topic “World cup 2014” ([T2, T4]). Meanwhile, the sentiments of “World cup 2014” are mainly concentrated within the range of content and elated, whereas sentiments for “Australian Politics” split into two parts with similar strength ([T3, T4]).

6 CONCLUSION

In this work, we introduced *SentimentClock* for visualizing the sentiments of time-varying microblog data. We illustrated our visualization design with two case studies to show its effectiveness in achieving various tasks related to sentiment analysis.

REFERENCES

- [1] M. M. Bradley and P. J. Lang. Affective norms for english words: Instruction manual and affective ratings. Technical report, 1999.
- [2] L. Feldman Barrett and J. A. Russell. Independence and bipolarity in the structure of current affect. *Journal of personality and social psychology*, 74(4):967, 1998.
- [3] M. C. Hao, C. Rohrdantz, H. Janetzko, D. A. Keim, U. Dayal, L. E. Haug, M. Hsu, and F. Stoffel. Visual sentiment analysis of customer feedback streams using geo-temporal term associations. *IV’13*, 2013.
- [4] S. Ramaswamy. Visualization of the sentiment of the tweets. *North Carolina State University*, Master Thesis, 2011.