

SentiCircles: A Platform for Contextual and Conceptual Sentiment Analysis

Hassan Saif¹, Maxim Bashevoy², Steve Taylor², Miriam Fernandez¹, and Harith Alani¹

¹ Knowledge Media Institute, Open University, UK

² IT Innovation, UK

Abstract. Sentiment analysis over social streams offers governments and organisations a fast and effective way to monitor the publics’ feelings towards policies, brands, business, etc. In this paper we present SentiCircles, a platform that captures feedback from social media conversations and applies contextual and conceptual sentiment analysis models to extract and summarise sentiment from these conversations. It provides a novel sentiment navigation design where contextual sentiment is captured and presented at term/entity level, enabling a better alignment of positive and negative sentiment to the nature of the public debate.

Keywords: social media, sentiment analysis

1 Introduction

With the emergence of social networking sites, fast and continuous streams of data are being generated by citizens about a variety of topics, including their opinions and arguments about policies, products, brands, etc. However, summarising and extracting sentiment from these large and continuous streams of data, as well as designing visualisation models to navigate such data, constitute difficult and important research problems.

Most current approaches for identifying the sentiment of posts can be categorised into one of two main groups: supervised approaches, such as [1],[3], which use a wide range of features and labelled data for training sentiment classifiers, and lexicon-based approaches, such as [8],[4], which make use of pre-built lexicons of words weighted with their sentiment orientations to determine the overall sentiment of a given text. Popularity of lexicon-based approaches is rapidly increasing since they require no training data, and hence they are more suited to a wider range of domains than supervised approaches [8]. Nevertheless, lexicon-based approaches have two main limitations. Firstly, the number of words in the lexicons is limited, which may constitute a problem when extracting sentiment from very dynamic environments such as the ones posed by social media, where new terms constantly emerge. Secondly and more importantly, sentiment lexicons tend to assign a fixed sentiment orientation and score to words, irrespective of how these words are used in the text. For example, the word “great” conveys different sentiment when associated with the word “problem” than with

the word “smile”. The sentiment that a word expresses is not static, but depends on the semantics of the word in the particular context it is being used. In addition, relying on the context only for detecting terms’ sentiment may also be insufficient. This is because the sentiment of a term may be conveyed via its conceptual semantics rather than by its context [2]. For example, the context of the word “Ebola” in “Ebola continues spreading in Africa!” does not indicate a clear sentiment for the word. However, “Ebola” is associated with the semantic type (concept) “Virus/Disease”, which suggests that the sentiment of “Ebola” is likely to be negative.

As part of our previous work we have been investigating novel sentiment analysis models that account for contextual and conceptual semantics in order to enhance the accuracy of existing lexical-based sentiment classification methods [7],[6],[5]. These models have been integrated into the SentiCircles platform and a novel interface has been designed to navigate social media conversations based on the sentiment computed by these models. More specifically, the objectives of the SentiCircles platform are: (i) apply contextual and conceptual sentiment analysis models for identifying the sentiment expressed in social media discussions and, (ii) provide visualisation mechanisms that enable a fine-grained summarisation and exploration of the computed sentiment.

Demo: A fully working platform will be demoed at the conference, running over a total of 12,000 posts. These posts span conversations around several topics related with the UK National policy on renewable energy targets for 2030.³ A tutorial of how to access and use the demo is available at:

<https://www.evernote.com/shard/s217/sh/8dba1de2-5353-4df2-89ce-2e4323a3eb36/36a950bb95b3deed>

2 The SentiCircles Sentiment Analysis Approach

The sentiment analysis component behind the SentiCircles platform uses the SentiCircle approach [5]. This approach accounts for contextual and conceptual semantics of words when computing sentiment. It detects the context of a term from its co-occurrence patterns with other terms in tweets. In particular, the context for each term t in a tweet collection \mathcal{T} is represented as a vector $\mathbf{c} = (c_1, c_2, \dots, c_n)$ of terms that occur with t in any tweet in \mathcal{T} . An example of this process is provided in Figure 2. Given a tweet collection \mathcal{T} , the target term m_{great} is represented as a vector $\mathbf{c}_{great} = (c_1, c_2, \dots, c_n)$ of terms co-occurring with term m in any tweet in \mathcal{T} (e.g., “pain”, “loss”, ..., “death”). The context vector \mathbf{c}_{great} is transformed into 2d circle representation. The center of the circle represents the target term m_{great} and points within the circle denote the context terms of m_{great} .

Conceptual semantics are incorporated into the approach by first extracting the entities from the posts (e.g., “Ebola”, “Africa”, and its conceptual types and subtypes (e.g., “Virus”, “Continent”) using AlchemyAPI⁴ and incorporating this information $\mathbf{s} = (s_1, s_2, \dots, s_m)$ within the contextual vector

³ <https://goo.gl/ykRDOA>

⁴ www.alchemyapi.com

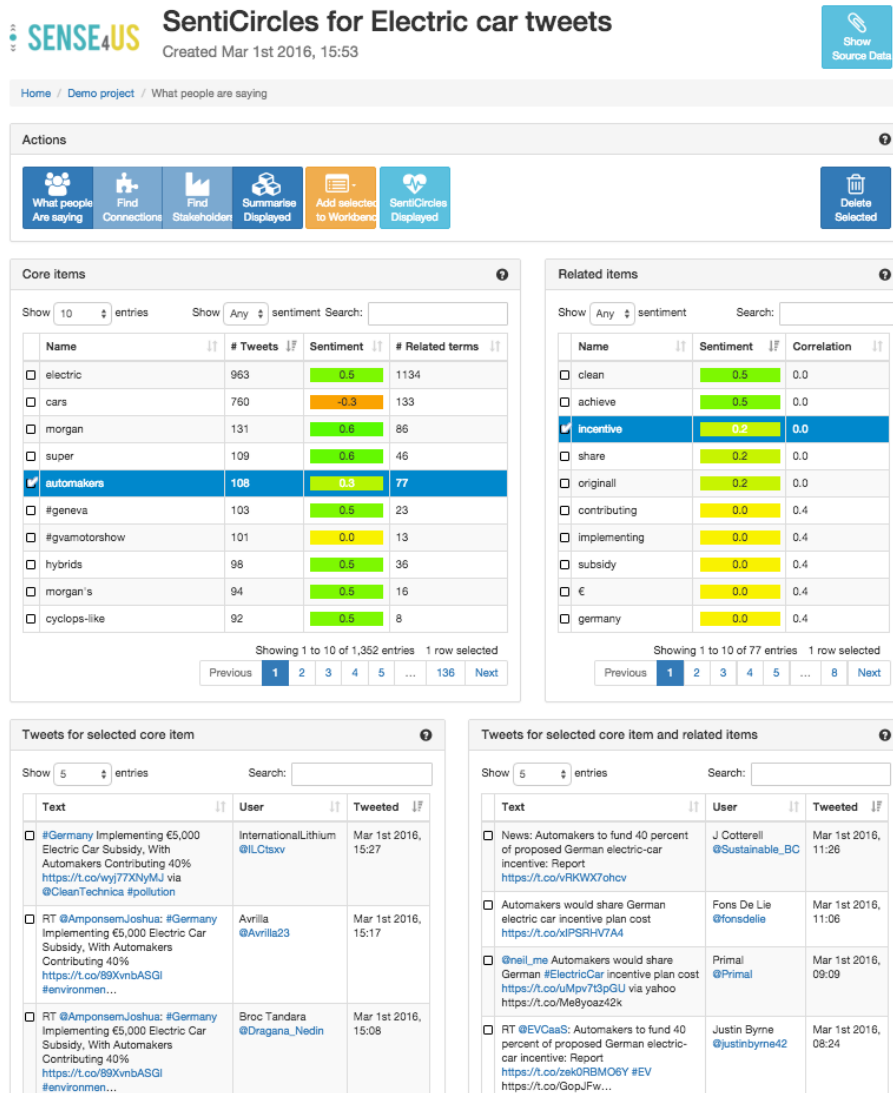


Fig. 1. SentiCircles visualisation for Twitter conversations around electric cars

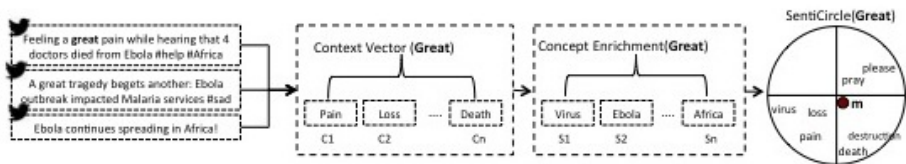


Fig. 2. Contextual and Conceptual sentiment of the word great

$\mathbf{c}_s = \mathbf{c} + \mathbf{s} = (c_1, c_2, \dots, c_n, s_1, s_2, \dots, s_m)$. The sentiment of t is then extracted by first transforming the term vector c_s into a 2d circle representation, and then extracting the geometric median of the points (context terms) within the circle. The position of the median within the circle represents the overall contextual sentiment of t . This simple technique has proven effective in calculating sentiment and entity as well as at tweet level (see [7],[6],[5] for more details)

3 The SentiCircles Sentiment Analysis Platform

In this section we present the designed visualisation for the SentiCircle Sentiment Analysis Platform. This visualisation has been designed to enable an easy navigation and exploration of the sentiment emerging from social media conversations. Figure 1 shows the sentiment emerging from a Twitter collection around electric cars. Each term emerging from the social media conversations is listed within the core items table (top left of the interface) alongside its corresponding contextual and conceptual semantics in the related items table (top right of the interface). These tables show the user to have a very quick overview of the issues people are discussing when talking about the topic in hand, along with their associated sentiment and relevant related information.

The core items table displays the terms and entities emerging from the social media conversations and for which sentiment has been computed, i.e., each core item is a SentiCircle (see Section 2), where the displayed term is at the center of the circle (e.g., electric, cars, automakers). Core items can be sorted and ranked according to: (i) the number of tweets in which they appear (i.e., how popular they are), (ii) their associated sentiment scores (from completely negative -1 to completely positive +1), and (iii) the number of related or contextual items associated to them. This information is represented in different columns of the core items table. Sorting capabilities are enabled on top of each column. For example, the sentiment column allows ranking the core items from more negative to more positive and vice-versa. Additionally, a range of colours is also provided to indicate positive (green), negative (red) and neutral (yellow) sentiment.

Contextual and conceptual semantics for each core item are presented in the related items table (top right of the figure) with their corresponding sentiment score, and the degree of correlation with respect to the core item. Examples of related items for the core item *automakers* include *clean*, *achieve*, *incentive*, etc. The degree of correlation represents how frequently do the co-occur in the post collection, which translates into the model on how strongly the related item influences the core item’s sentiment. The top of each column enables to sort the rows based on their numeric value.

The tables at the bottom part of the interface show examples of posts in which the selected core item appears (left bottom table), and examples of posts where both, the core item and the selected related item appear (right bottom table). The presented model and its visualisation enables the user to navigate social media conversations, observe what are the core emergent items around the topic at hand, what is the sentiment perceived towards those items, and what are the reasons (context) behind it.

4 Feedback and Future Work

The SentiCircles platform is part of the Sense4us project⁵, a project focused on the development of tools that can support better policy making. The platform was showcased to 16 Members of Parliament (MPs) from the State Parliament of North Rhine Westphalia, the German Bundestag and the European Parliament. It was very well received as a tool that could enhance the collection of feedback, and speeding up the reaction to any concerns or challenges raised by citizens. MPs highlighted that positive and negative sentiment appeared better aligned to the nature of the public policy debate than in other tools, since SentiCircles shows them the key items under discussion within the conversations and allows them to investigate why sentiment is negative or positive by navigating between the core and related items. Transparency was raised as an important issue, since MPs need to be aware of the limitations, particularly in terms of data, model and users, when using social media analysis tools to inform policy making. We are currently working on preparing documentation, as well as various interface modifications to enhance the transparency of the results obtained by the tool.

5 Conclusions

This paper describes the SentiCircles platform; a web based tool for assessing and monitoring sentiment in public social media. The platform applies contextual and conceptual sentiment analysis models to extract and summarise sentiment. Positive feedback has been received so far about the tool by several MPs when tested in a policy making context.

References

1. L. Barbosa and J. Feng. Robust sentiment detection on twitter from biased and noisy data. In *Proceedings of COLING*, Beijing, China, 2010.
2. E. Cambria. An introduction to concept-level sentiment analysis. In *Advances in Soft Computing and Its Applications*, pages 478–483. Springer, 2013.
3. E. Kouloumpis, T. Wilson, and J. Moore. Twitter sentiment analysis: The good the bad and the omg! In *Proceedings of the ICWSM*, Barcelona, Spain, 2011.
4. B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, 11:122–129, 2010.
5. H. Saif, M. Fernandez, Y. He, and H. Alani. SentiCircles for contextual and conceptual semantic sentiment analysis of twitter. In *Proc. 11th Extended Semantic Web Conf. (ESWC)*, Crete, Greece, 2014.
6. H. Saif, Y. He, and H. Alani. Semantic sentiment analysis of twitter. In *Proc. 11th Int. Semantic Web Conf. (ISWC)*, Boston, MA, 2012.
7. H. Saif, Y. He, M. Fernandez, and H. Alani. Contextual semantics for sentiment analysis of twitter. *Information Processing & Management*, 2015.
8. M. Thelwall, K. Buckley, and G. Paltoglou. Sentiment strength detection for the social web. *J. American Society for Information Science and Technology*, 63(1):163–173, 2012.

⁵ <http://www.sense4us.eu/>