

# SENTIMENT ANALYSIS OF SOCIAL MEDIA TEXTS: A SEMEVAL PERSPECTIVE

**Preslav Nakov** (preslav.nakov@gmail.com)

Qatar Computing Research Institute, Qatar

Internet has democratized content creation leading to the rise of social media and to an explosion in the availability of short informal text messages. Microblogs such as Twitter, weblogs such as LiveJournal, social networks such as Facebook, and instant messengers such as Skype and Whatsapp are now commonly used to share thoughts and opinions about anything in the surrounding world. This proliferation of social media content has created new opportunities for studying public opinion, with Twitter being especially popular for research purposes due to its scale, representativeness, variety of topics discussed, as well as ease of public access to its messages.

I will present the evolution of a semantic analysis task that lies at the intersection of two very trendy lines of research in contemporary computational linguistics: (i) sentiment analysis, and (ii) natural language processing of social media text.

The task was part of SemEval (the International Workshop on Semantic Evaluation, a semantic evaluation forum previously known as SensEval), and it ran in 2013, 2014, and 2015, attracting the highest number of participating teams at SemEval in all three years; there is an upcoming edition in 2016. The task included the creation of a large contextual and message-level polarity corpus consisting of tweets, SMS messages, LiveJournal messages, and a special test set of sarcastic tweets. The evaluation attracted 44 teams in 2013, 46 in 2014, and 41 in 2015, which used a variety of approaches. The best teams were able to outperform several baselines by sizable margins with improvements over the years.

The task has fostered the creation of some freely-available, and by now widely used, resources such as NRC's Hashtag Sentiment lexicon and the Sentiment140 lexicon, which the NRC-Canada team initially developed for their participation in SemEval-2013 task 2, and which were key for their winning the competition.

The 2015 and 2016 editions of the task switched focus to sentiment with respect to a topic, on a positive/negative/neutral (2015) or on a five-point scale (2016). The latter is used for human review ratings on popular websites such as Amazon, TripAdvisor, Yelp, etc. From a research perspective, moving to an ordered five-point scale means moving from classification to ordinal regression.

Another shift in 2015 was from sentiment towards a topic in a single tweet to sentiment towards a topic in a set of tweets (trend detection). This represents a move from classification to quantification. In real-world applications, the focus often is not on the sentiment of a particular tweet, but rather on the percentage of tweets that are positive/negative. In 2016, trend detection will be also offered on a five-point scale, which gets us even closer to what business (e.g., marketing studies), and researchers, (e.g., in political science or public policy), want nowadays. From a research perspective, this is a problem of ordinal quantification.

I will introduce the above subtasks in detail, and I will explain the process of creating the training and the testing datasets. I will briefly describe some of the participating systems, and the overall results, in comparison with a number of baselines. Special attention will be paid to the lessons learned with analysis across a number of dimensions such as progress over the years, performance on out-of-domain data, utility of context in the contextual polarity task, impact of training data size, of using external resources, of negation handling, etc. Finally, I will compare the task to other related tasks at SemEval and beyond.