SENTIMENT ANALYSIS: APPLICATION, CHALLENGES, TOOLS

Oksana Semyda

PhD., Associate Professor,

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

In recent years we have seen a rising interest in the study of subjectivity, opinion and evaluation. This has opened up the field of sentiment analysis at the intersection of linguistics and computer science. A new field of knowledge emerged, known as sentiment analysis (SA). SA is an opinion mining or emotion detection. It is employed to analyze different samples of texts with the purpose to find out the emotional tone, connotation and implicit meanings of the texts. The need for sentiment analysis lies in the necessity to process, systematize and analyze the texts from the Internet, social media, customers feedback. Politicians, industrial companies, sociologists want to know the tendencies prevailing in modern society so they are interested in mining/studying people's opinions that can be revealed from the texts created by them.

The vigorous information exchange on the Internet led to the emergence of the SA. This tool became widely used for collecting and analyzing attitudes and content on the Internet and social networks. It deals with responses and clients' feedback to find out their acceptance or rejection of a product. Knowledge about customers' attitudes and preferences can help to improve the company's sales and therefore profits. Opinions and attitudes can be revealed from social networking sites, and those can be used by politicians, psychologists and business owners to implement the best decisions (Ameen, Manjula & Naik, 2019).

SA represents an important element for many spheres of our life: politics, business, production as it helps predict financial performance, understand consumers' perception, provide early warnings, define election outcomes etc. Scientists name 5 stages in data sentiment analysis. These stages include data collection, text preparation, sentiment detection and classification and output representation (D'Andrea, Ferri, Grifoni & Guzzo, 2015).

By the data collection we imply collecting data from blogs, forums, social networks.

To analyze this data manually is impossible as it is disorganized and nonuniform. Text preparation includes data extraction where irrelevant non-textual data is eliminated. Then comes the stage of sentiment detection itself, that is sentences with reviews and opinions are analyzed and objective information is rejected.

In the sentiment classification step, the sentences are classified concerning a positive or negative meaning, likes or dislikes.

The final stage of the SA is output presentation which is always in the form of graphs, pie charts or images. It can illustrate the results taking into account a timeline, showing changes with time.

The table below demonstrates generalized sentiment analysis applications in various social spheres (D'Andrea et al., 2015).

Business	Consumers voice
	Brand reputation
	Online advertising: Blogger Centric Contextual Advertising Dissatisfaction oriented online advertising On-line commerce
Politics	Voting advise applications
	Clarification of politicians' positions
Publications	Real-world events monitoring
	Legal matters "blawgs"
	Policy or government-regulation proposals
Finance	Prices of commodities and shares evolution
	Financial risk individuation

Table 1. Sentiment Analysis Application

Scientists outline the main three approaches to SA:

- machine learning
- lexicon-based approach
- hybrid approach

The machine learning approach helps predict the polarity of sentiments based on learned data and information. Here a classifier is built that can determine the polarity of new texts. This classifier can be taught and extended. Machine learning is regarded as accurate but among disadvantages, we can mention that software tools are trained very specific data that is not transferable.

The lexicon based approach uses a predefined list of words and does not require any prior training. In this list, every word is marked with a specific sentiment. The hybrid approach is the combination of the two previously mentioned.

Also we need to mention some tools used for SA. They are quite expensive but they can be ranged from available online to paid advanced software tools or even specially designed for you. Thorough text analysis can be provided by using sentiment analysis tools that employ machine learning and processing of natural language. The main idea is that the more online samples you analyze, the more accurate results you get. The most common and popular tools are Brand24, Clarabridge, Repustate, OpenText, Emoticons, LIWC, SentiStrengh, Senti WordNet, SenticNet, Happiness Index, AFINN, PANAS-t, Sentiment140, NRC, EWGA and FRN.

SA presents such challenges. The matter is that people express their opinions in a complex way using sarcasm, ironies etc, that is, the misleading content. In his research Mohammad (2015) discussing the challenges gives such an example: “How can someone sit through the movie” is an extremely negative sentiment yet contains no negative lexicographic word. Other factors that create challenges are a speaker’s emotional state, the dependence of evaluation on the trends in a certain society, external and internal politics of a country, quotes and retweets.

To conclude, we can say that the field of SA has urged great interest among researchers. Although it is a new field, a lot of work has been done to discover it. Scientists continue working on new methods, models and tools for opinion mining as well as on the accuracy of the results and new ways of applications.

References

Ameen, A. Q. A., Manjula, B., & Naik, R. L. (2019). A Study of sentiment analysis: Concepts, techniques, and challenges. *Proceedings of International*

Conference on Computational Intelligence and Data Engineering, Lecture Notes on Data Engineering and Communications Technologies, 28, 147-162. https://doi.org/10.1007/978-981-13-6459-4_16

D'Andrea, A., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications*, 125(3), 26-33. Retrieved from https://www.ijcaonline.org/research/volume125/number3/dandrea-2015-ijca-905866.pdf

Mohammad, S. M. (2015). Challenges in Sentiment Analysis. *A Practical Guide to Sentiment Analysis*, 2-23. https://doi.org/10.1007/978-3-319-55394-8_4