**Levitating Sentimental Analysis**

**Overview and Introduction**

Sentiment analysis (also known as opinion mining or emotion AI) is the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to the voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine.

There are three main classification levels in SA: document-level, sentence-level, and aspect-level SA. Document-level SA aims to classify an opinion document as expressing a positive or negative opinion or sentiment. It considers the whole document a basic information unit (talking about one topic). Sentence-level SA aims to classify sentiment expressed in each sentence. The first step is to identify whether the sentence is subjective or objective. If the sentence is subjective, Sentence-level SA will determine whether the sentence expresses positive or negative opinions. Wilson et al. have pointed out that sentiment expressions are not necessarily subjective in nature. However, there is no fundamental difference between document and sentence level classifications because sentences are just short documents. Classifying text at the document level or at the sentence level does not provide the necessary detail needed opinions on all aspects of the entity which is needed in many applications, to obtain these details; we need to go to the aspect level. Aspect-level SA aims to classify the sentiment with respect to the specific aspects of entities. The first step is to identify the entities and their aspects. The opinion holders can give different opinions for different aspects of the same entity like this sentence “The voice quality of this phone is not good, but the battery life is long”. This survey tackles the first two kinds of SA.

The data sets used in SA are an important issue in this field. The main sources of data are from the product reviews. These reviews are important to the business holders as they can take business decisions according to the analysis results of users’ opinions about their products. The review sources are mainly reviewed sites. SA is not only applied to product reviews but can also be applied to stock markets, news articles, or political debates. In political debates, for example, we could figure out people’s opinions on certain election candidates or political parties. The election results can also be predicted from political posts. Social network sites and micro-blogging sites are considered a very good source of information because people share and discuss their opinions about a certain topic freely. They are also used as data sources in the SA process.

Sentiment Classification techniques can be roughly divided into machine learning approach, lexicon-based approach, and hybrid approach. The Machine Learning Approach (ML) applies the famous ML algorithms and uses linguistic features. The Lexicon-based Approach relies on a sentiment lexicon, a collection of known and precompiled sentiment terms. It is divided into the dictionary-based approach and corpus-based approach which use statistical or semantic methods to find sentiment polarity. The hybrid Approach combines both approaches and is very common with sentiment lexicons playing a key role in the majority of methods.

The text classification methods using the ML approach can be roughly divided into supervised and unsupervised learning methods. The supervised methods make use of a large number of labeled training documents. The unsupervised methods are used when it is difficult to find these labeled training documents.

**Conclusion and future work**

This survey paper presented an overview of the recent updates in SA algorithms and applications. Fifty-four of the recently published and cited articles were categorized and summarized. These articles give contributions to many SA-related fields that use SA techniques for various real-world applications. After analyzing these articles, it is clear that the enhancements of SC and FS algorithms are still an open field for research. Naïve Bayes and Support Vector Machines are the most frequently used ML algorithms for solving SC problems. They are considered a reference model where many proposed algorithms are compared to.

The interest in languages other than English in this field is growing as there is still a lack of resources and researches concerning these languages. The most common lexicon source used is WordNet which exists in languages other than English. Building resources, used in SA tasks, are still needed for many natural languages.

Information from micro-blogs, blogs, and forums as well as a news source, is widely used in SA recently. This media information plays a great role in expressing people’s feelings, or opinions about a certain topic or product. Using social network sites and micro-blogging sites as a source of data still needs deeper analysis. There are some benchmark data sets especially in reviews like IMDB which are used for algorithms evaluation.

In many applications, it is important to consider the context of the text and the user preferences. That is why we need to make more research on context-based SA. Using TL techniques, we can use related data to the domain in question as training data. Using NLP tools to reinforce the SA process has attracted researchers recently and still needs some enhancements.

**Bibliography**

* <https://www.sciencedirect.com/science/article/pii/S2090447914000550>
* <https://en.wikipedia.org/wiki/Sentiment_analysis>