

Theoretical background to the machine learning approach of sentiment analysis



**ASSIGN
BUSTER**

Automatic sentiment analysis regards the extraction of a sentiment from an unstructured source such as text, images or audio. The recognized sentiments can be classified as positive or negative, or a more fine grained sentiment classification scheme can be used.

Sentiment analysis of text, also called opinion mining, only recently received a large interest from the academic community and commercial companies. What people write on persons, products or institutions has an important value in our society and the World Wide Web is an excellent source of such information.

The automatic analysis of sentiments on data found on the World Wide Web is useful for any company or institution caring about quality control. For the moment, getting user feedback means bothering him or her with surveys on every aspect the company is interested in. Making a survey for each product or feature, designing the format, distribution and timing of the survey (sending a form right after purchase might not be very informative), and the reliance on the goodwill of people to take the survey are expensive and time-consuming tasks, yielding not always accurate results. Surveying by means of questionnaires can be made obsolete by gathering such information automatically from the World Wide Web. One of the sources are blogs (short for "web logs"), a medium through which the blog owner makes commentaries about a certain subject or talks about his or her personal experiences, inviting readers to provide their own comments. Other sources are customer review sites and electronic discussion boards or forums, where people can discuss all kinds of topics, or ask for other people's opinions.

There are several additional advantages to automated sentiment analysis. First, the people who share their views usually have more pronounced opinions than average. These opinions are additionally influencing others reading them, leading to so-called word-of-mouth marketing. Extracting these opinions is thus extra valuable. Second, opinions are extracted in real-time, allowing for quicker response times to market changes and for detailed time-based statistics that make it possible to plot trends over time. Last, but not least in information retrieval opinion mining assists in discriminating opinionated documents from documents that present the information in a neutral way.