

SENTIMENT ANALYSIS

Shubhashree Dash

USCID : 4328734481

shubhash@usc.edu

Natural Language can be decomposed into three parts such as syntax, semantics and pragmatics. The basic stages of natural language processing, it analyses the order or structure of the sentence, the literal meaning of the sentence, and then the meaning in context. This three stage analysis is a good starting point, however for the current need, a finer decomposition is required. The stages are tokenization, lexical analysis, syntactic analysis, semantic analysis and then pragmatic analysis. Sentiment analysis can be covered under the semantic analysis of text.

Textual information can be categorised into two main types : facts and opinions. Facts are ground truths, objective expressions about an event, entity, or their property. Opinions on the other hand are personal, derived from personal experience. others' experiences and opinions are important for people to make decisions. The World Wide Web is now filled with user-generated content in the form of forums, blogs, user reviews. These content is now used by users to chose whether they should buy a particular product or service. The user now are not limited to the opinions of their friends and families, but also consider the opinions of strangers on the internet. This data is humongous and it is difficult for a human to go through the varied form of content, and identify the parts which can identify the option of the user for a particular product. This forms the need of automatic detection and summarisation systems. Sentiment analysis, also known as opinion mining comes to aid to solve this problem.

Sentiment Classification

Sentiment classification is a key topic in sentiment analysis research. In this, an opinionated text is classified into either positive or negative opinion. The classes are not limited to positive or negative, but can be extended to neutral or use a point based system like the Amazon 5 star system to review products.

There are five specific problems within the field of sentiment analysis. Document-level, Sentence-level, Aspect-based, Comparative, Sentiment lexicon acquisition.

Document level sentiment analysis : A single document, for example a review of a single product on Amazon, can have a single opinion which expresses where the author liked or disliked the product. The task is to find a single opinion expressed in the document, by an author. This assumes that a document has an opinion on a single product. The opinion is from a single author. The opinion is expressed in the entire document.

Sentence level sentiment analysis : A single document can contain multiple opinions, for example blogs posts that describe the latests products, author can express things they liked or didn't like about the products. The task here is to determine if the sentence is subjective or objective, and for subjective sentence find the opinion expressed in a single sentence of the content.

Aspect level sentiment analysis (feature based sentiment analysis) : For products with multiple attributes, authors can have different opinions about each attribute of the product. The task is to identify all the opinions present in the document along with the corresponding aspect of the product. It can also be viewed as an information extraction problem.

Comparative sentiment analysis : Users sometimes compare a product to another to express their opinions. They compare two products, instead of giving a direct opinion.

For example, “iPhone camera is better than my previous Galaxy” instead of “iPhone has a great camera”. There are two main types of comparison: comparative, and superlative, as well as increasing and decreasing. The task is to identify comparative sentences and categorise them into types/classes.

Sentiment Lexicon Acquisition : Lexicons can be used to identify opinions on a product in a sentence. It uses opinion words and phrases to identify whether the opinions are positive or negative. For example, in “Kindle Fire has an amazing display,” the adjective ‘amazing’ modifies the noun ‘display,’ so given that ‘amazing’ is a sentiment expression which is positive in nature. This expression can be used to come to the conclusion that the author has a positive opinion about the display.

Sentiment Classification Techniques

Sentiment Classification Techniques can be divided into three approaches: Machine Learning, Lexicon Based, and Hybrid.

Machine learning approach uses linguistic features in some machine learning models. This can be further divided into supervised and unsupervised learning depending on the corpus. Supervised learning methods use a dataset which is annotated, and unsupervised methods are used when the dataset is difficult to label. Machine learning methods use syntactic or linguistic features to classify text. Here are some classifiers which assign class based on the opinion expressed:

Naive Bayes : It calculates the posterior probability of a class based on the distribution of words in the document. It uses Bayes theorem to predict the probability of class given the set of words or features from a document.

Bayesian Networks : It assumes that all the features of the document are dependent on each other and computes a complete joint probability distribution table for all variables. This is computationally expensive thus rarely used.

Support Vector Machines : It determines a vector that separates the document vectors pertaining to different classes. It tries to maximise the margin between the classes. It is suitable for text as the text data is sparse and features are generally correlated to one other.

Neural Networks : It uses many neurons to classify the text document. The input to the neuron is a vector per document. Each vector consists of word frequency per document.

Decision Tree Classifier : It uses presence or absence of a phrase or word as a condition to decompose the training dataset into classes.

Lexicon based approach uses sentiment lexicon, which are previously known and compiled. It is further divided into dictionary based - which finds opinion seed words from the document and then finds its synonym and antonym in the dictionary, and corpus based - which uses a seed list and then finds the opinion words in the dataset

The hybrid approach combines both the machine learning and lexicon based approaches.

Applications

Sentiment analysis has many applications in the real world. Most common application is summarisation of review of products. They are also used to monitor brand reputation on blogging and micro-blogging sites. Sentiment analysis can also help in political campaign

to identify voter alignment as well as their views on different topics. Sentiment analysis can also be used as an addition to other systems, such as a recommendation system. It can be used to avoid recommending products which have lot of negative reviews.

References

Anna Atefeh Farzindar; Diana Inkpen, Natural Language Processing for Social Media: Third Edition , Morgan & Claypool, 2020.

Liu, Bing (2010). "Sentiment Analysis and Subjectivity" (PDF). In Indurkha, N.; Damerau, F. J. (eds.). Handbook of Natural Language Processing (Second ed.).

Dale, Robert (2010). "Classical Approaches to Natural Language Processing" (PDF). In Indurkha, N.; Damerau, F. J. (eds.). Handbook of Natural Language Processing (Second ed.).

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113.

Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82.

Tripathy, A., Agrawal, A., & Rath, S. K. (2015). Classification of Sentimental Reviews Using Machine Learning Techniques. Procedia Computer Science, 57, 821–829.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135.