

Sentiment Classification for Information Retrieval

With so many opinions on the web about products and services it is becoming harder for people to evaluate what the majority opinions about the products and services is. Sentiment classification is an important approach for evaluation in sentiment mining. Also known as affect or polarity classification, the approach calls to address the problem of understanding opinions by providing aggregate views and segmenting the text into classes.

Sentiment classification is commonly used to evaluate documents for determining a positive or negative opinion or sentiment on an object. The opinion words are very important in this approach. Words like great, excellent, horrible, bad etc that indicate a positive or negative opinion are the key startup points in sentiment classification.

The researches in sentiment mining have suggested different approaches to analyze classifiers. One of the popular techniques is the star rating as polarity signals. Many websites on the internet provide ratings about products and services to give an overall perception about them. But there are still many posts, forums and reviews on the internet that do not provide explicit ratings. Sentiment classification has been widely used for natural language processing (NLP) communities.

Approaches in Sentiment Classification

Various text categorizations algorithms are being used for sentiment classification. The machine learning algorithms like EM-based Naïve Bayes. Maximum Entropy and Support Vector Machines are popular methods to process information. Naïve Bayes works well on text categorization by assuming that the presence or absence of a particular feature is unrelated to the presence and absence of other features.

Maximum Entropy is also a popular approach to sentiment classification which is also feature based approach. It is used widely with natural language processing. It is similar to using logistic regression to find a distribution over the classes. Support Vector Machine is useful in identification of features through vector distribution. It is mostly used for feature processing in tweets and blogs.

One of the most common approaches used in text classification is keyword based classification. In this type of classification the terms and fixed expressions are used as sentiment indicators. These terms can mostly be adjectives like awesome or awful and fixed expressions could be a combination of two or more words like stay away or dreams come true. These words can be prepared manually or through machine learning algorithms whichever provides the best result. This approach is mostly compared with sentiment classification for effectivity.

The main issue with sentiment classification is that while evaluating documents it makes the assumptions that each document to be evaluated focuses on a single object and contains opinions by an opinion holder. This assumption is used to determine the semantic orientation of

the opinion expressed for an object. However this approach can be limited when applied to demanding documents. For documents of reviews containing text of negative stance with positive terms the sentiment classification can be difficult.

Linguistic Dimensions to Sentiment Classification

Adding linguistic dimension to the sentiment classification in sentiment mining can provide effective results. There are two contributive enhancements in this case. Firstly the addition of composite features that are of syntactic nature can set tentative classifiers. Verb objects pairs are one example in this regard. Secondly using polarity markers will help for extracting the information required in order to be able to use it properly. Polar words are important to the overall polarity of the context.

Sentiment classification is an important part of sentiment mining and a lot of research is still being done to use classification methods on different types of text for efficient sentiment detection.