

Evaluating Sentiment Mining

Since sentiment mining or opinion mining relates to the attitude, opinion and emotional state of a writer or speaker the evaluation methodologies are slightly complicated. It is difficult to quantify sentiments. Sentiment mining evaluations rely heavily on context. Some evaluations might only assume subjectivity and polarity and some work might only include information retrieval based on relevance and subjectivity and not polarity.

There are degrees of sentiments to be measured. Almost all text can contain some amount of sentiment. This can be a debatable stance but the some kind of gradation of sentiments depicting either the presence or absence of sentiments is always present. Sentiments can also be measured at different granularities. It can be measured at a paragraph level, document level, sentence level, phrase level, pattern level or a combination of these levels in a given content.

In any form of evaluation for sentiment mining the objective of measurement is very important. Out of the families of evaluation three known types of evaluation methods include document level evaluation, aggregate level evaluation and information retrieval evaluation.

Document Level Evaluation

This is the most common way to evaluate sentiment mining specially in social media analytics. This is also known an obvious metric for analysts who are working on text classification models for document level accuracy.

In terms of analyzing a social media monitoring solutions the set of documents are read and the agreements are marked. There are labels and classes that are assigned to a document. The approach requires agreeing on a consistent set of labels and classes by assigning them on the document. Classes or labels could simply be positive, negative or neutral. Other statistical measures may also be used. For document level evaluations the following things are important

1. The number of classes matter a lot in this type. It is also important to compare similar classifiers for accuracy.
2. The distribution of class label must be accurate. Random guessing with equal class probabilities should get the same true distribution as assigned.
3. Population of documents determines the level of difficulty for this type of evaluation for example it can be easy to classify a tweet than a blog post.

Aggregate Level Evaluation

The primary use of opinion mining or sentiment mining is tracking the attitudes about a topic or opinion of a population. It is important to consider the following points for sentiment evaluation according to aggregate level distribution.

Firstly it is important to directly assess real versus estimated distribution for making right conclusions. Using statistically grounded measures like a chi-squared test or a transparent metric like error or distance brings more closer to true conclusions.

Secondly to get a better idea of performance it is important to deconstruct the problem. Measuring how well the solution does on subjectivity, relevance and polarity is better than looking at all classes at once.

Lastly when comparing claims, it is important that they are on the same number of class labels, type of data (like blog or forum) and same topic (like true distribution).

Information Retrieval Evaluation

This is also an important evaluation method. When reading a negative post it does not have to be labeled with discrete titles like positive, negative and neutral. A degree of belonging can be used here. An individual post can be labeled with a single class having varied sentiments. Rank ordering metric can be applied here. Evaluating sentiments according to the ability to retrieve appropriate examples needs the following points to be considered.

1. Avoiding micro level averaging precisions and looking at each of the class directly.
2. Like other evaluation methods number of classes and their distribution matters to give the right precisions on the type of classes.
3. When comparing claims it must be checked that they are on the same number of class labels, topic and type of data.