



Unstructured Data Analytics

Octillion Technology Inc.

Making the right decision at the right time is the only way to succeed. Executives and managers in private enterprises or in government departments need real-time and accurate information to make the right decisions. Octillion Technology is committed to providing advanced decision making solutions to its clients.

Unstructured data analytics refers to analyzing text data. Analysis of textual data can be used in determining customers' attitudes with respect to products, services or policies, and in tracking these attitudes or opinions as they change over time. They can also be used to gauge opinions of the people to predict effectiveness of campaigns run by industries, governments or other private agencies.

Detecting attitudes, opinions and sentiments applies to any database or stream of textual information, especially user-generated content, such as corporate reports, blogs, web pages, bulletin boards, emails, chat rooms, etc. The process for detecting attitudes and sentiments requires three basic steps:

1. Collecting a training corpus of texts, human annotated as to their sentiment
2. Building a set of patterns associated with positive, negative and neutral sentiments
3. Training a statistical machine learning system possibly augmented with a suite of interpretation rules that can classify new texts into the desired categories.

Most previous work in sentiment detection has skipped the second step and essentially used words instead of patterns. For instance, words such as "great", "desirable", "better" are associated with positive sentiments, and words such as "awful", "terribly", "dangerous" have been associated with negative sentiments. However, this simplistic approach, avoiding language analysis, has proven unreliable. A text can say that a policy is "not at all desirable" (negative sentiment), or a product is "terribly good" (positive sentiment), which cannot be predicted from simple word lists. Our approach uses both patterns and words (the former with greater weight) to improve performance.

We also differ from earlier approaches in terms of measuring possible levels of sentiment (e.g. strongly negative vs. weakly negative), and in terms of tracking changes over time (e.g. opinion is gradually shifting towards the positive).

We have also found that for practical applications we need to combine sentiment detection with topic detection. If a stream of texts contains mixed attitudes, it could be that the positive ones refer to one topic (or product or policy) and the negative ones about a different one. Topic detection is a well-studied field also addressed by statistical learning methods, with which Carnegie Mellon University and Octillion Technology are very familiar.

Octillion is a technology company with a difference. We have the expertise to provide solutions with the back-end IT infrastructure and user interfaces. Our engineers use innovative and leading technologies to develop these solutions.