

# Sentiment Analysis: A Study

Aniket Redkar<sup>1</sup>, Sagar Lakhani<sup>2</sup>, Vinit Parekh<sup>3</sup>, Gautam Gala<sup>4</sup> and Prof. Snehal Kathale<sup>5</sup>

Department of Information Technology, Athana College of Engineering

## ABSTRACT

*Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis is a type of natural language processing for tracking the mood or opinion of the public about a particular product or topic or service. Our goal in this survey has been to cover techniques and approaches that promise to directly enable opinion-oriented information-seeking systems, and to convey to the reader a sense of our excitement about the intellectual richness and breadth of the area.*

**Keywords:** Sentiment Analysis, Natural Language Processing, Opinion Identification, Machine Learning

## 1. INTRODUCTION

Sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation, affective state or the intended emotional communication. A basic task in sentiment analysis classifying the polarity of a given text at the document, sentence, or feature/aspect level — whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. Advanced, "beyond polarity" sentiment classification looks, for instance, at emotional states such as "angry," "sad," and "happy."

Sentiment analysis concentrates on attitudes, whereas traditional text mining focuses on the analysis of facts. There are few main fields of research predominate in Sentiment analysis: sentiment classification, feature based Sentiment classification and opinion summarization. Sentiment classification deals with classifying entire documents according to the opinions towards certain objects. Most of our decisions in the real world are influenced by us thinking how other people will perceive our decision. In simple words, "What others will think if I do this" has always occupied our mind, may be just for few seconds, while making important decisions.

Just a decade ago, when internet was not so popular, people used to take decisions on buying some service or product based on their friends or critics recommendation. However, the amount of information available for decision making was limited. But with the popularity of internet and the big-data explosion, tremendous amount of information is available, which can be used to objectively make important decisions. Today, more and more people have started using internet to make decisions. These decisions include seeking an opinion on brands, products, services, religion, politics, economics, entertainment etc. Remainder of the Paper is divided as Section 2 which deals with Methodologies involved in Analysing Data, Section 3 deals with various Sources of data that can be worked on and the concluding Section 4 refers to Application of Sentiment Analysis in Daily Activities.

## 2. METHODOLOGY

### 2.1 Formation of Problem

Motivated by different real-world applications, researchers have considered a wide range of problems over a variety of different types of corpora. We now examine the key concepts involved in these problems. This discussion also serves as a loose grouping of the major problems, where each group consists of problems that are suitable for similar treatment as learning tasks.

### 2.2 Sentiment polarity and degrees of positivity

One set of problems share the following general character: given an opinionated piece of text, wherein it is assumed that the overall opinion in it is about one single issue or item, classify the opinion as falling under one of two opposing sentiment polarities, or locate its position on the continuum between these two polarities. A large portion of work in sentiment-related classification/regression/ranking falls within this category. The binary classification task of labeling an opinionated document as expressing either an overall positive or an overall negative opinion is called sentiment polarity classification or polarity classification. Although this binary decision task has also been termed sentiment classification in the literature, as mentioned above, in this survey we will use "sentiment classification" to refer broadly to binary categorization, multi-class categorization, regression, and/or ranking.

### 2.3 Opinion Identification

Work in polarity classification often assumes the incoming documents to be opinionated. For many applications, though, we may need to decide whether a given document contains subjective information or not, or identify which portions of the document are subjective. At least one opinion-tracking system rates subjectivity and sentiment separately. The problem of distinguishing subjective versus objective instances has often proved to be more difficult than subsequent polarity classification, so improvements in subjectivity classification promise to positively impact sentiment classification

### 2.4 Perspective

Much work on analyzing sentiment and opinions in politically-oriented text focuses on general attitudes expressed through texts that are not necessarily targeted at a particular issue or narrow subject. The other point of departure from the polarity classification problem is that the labels being considered are more about attitudes that do not naturally correspond with degree of positivity. While assigning simple labels remains a classification problem, if we move farther away and aim at serving more expressive and open-ended opinions to the user, we need to solve extraction problems.

## 3. RESOURCES

Users opinion is a major criterion for the improvement of the quality of services rendered and enhancement of the deliverables. Blogs, review sites, data and micro-blogs provide a good understanding of the reception level of the products and services

### 3.1 Micro-Blogging

Twitter is a social networking and micro-blogging service that allows users to post real time messages, called tweets. Tweets are short messages, restricted to 140 characters in length.

Blog 06

The University of Glasgow distributes this 25GB TREC test collection, consisting of blog posts over a range of topics. Access information is available at [http://ir.dcs.gla.ac.uk/test\\_collections/access\\_to\\_data.html](http://ir.dcs.gla.ac.uk/test_collections/access_to_data.html). Included in the data set are “top blogs” that were provided by Nielsen

Buzz Metrics and “supplemented by the University of Amsterdam”, and some spam blogs, also known as “sp logs”, that were planted in the corpus in order to simulate a more realistic setting. Assessments include relevance judgments and labels as to whether posts contain relevant opinions and what the polarity of the opinions was (positive, negative, or a mixture of both).

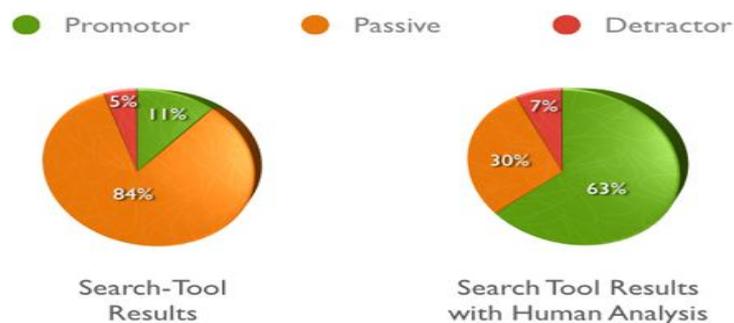
### 3.2 Reviews

A host of resources are available over the internet for data sources some of these can be mentioned as

Review Datasets: dataset available is <http://www.cs.uic.edu/liub/FBS/CustomerReviewData.zip>. This dataset consists of reviews of five electronics products downloaded from Amazon and Cnet (Hu and Liu ,2006; Konig & Brill ,2006 ; Long Sheng ,2011; Zhu Jian ,2010 ; Pang and Lee ,2004; Bai et al. ,2005; Kennedy and Inkpen ,2006; Zhou and Chaovalit ,2008; Yulan He 2010; Rudy Prabowo ,2009; Rui Xia ,2011).

## 4. APPLICATIONS

### SENTIMENT360 The Human Advantage



Comparison of automated sentiment results before and after human analysis.  
Study performed for a major TV network, March 2010

Figure 1: Human vs. Machine Sentiment Analysis

The applications for sentiment analysis are endless. Nowadays, the use of Sentiment Analysis is widely seen in social media monitoring, review sites, various forums, etc and is seen as quickest way to analyze public reactions towards a particular product. However, it is also practical for use in business analytics and situations in which text needs to be analyzed. Thousands of text documents can be processed for sentiment (and other features including named entities, topics, themes, etc.) in seconds, compared to the hours it would take a team of people to manually complete. Fig4.1 shows a study performed by Sentiment360 for a major TV network. In this, we can see the differences in the accuracy of sentiment prediction between human analysis and a tool based analysis. Hence tool based analysis has proved to be more accurate.

## **5. CONCLUSION AND FUTURE WORK**

In future, more work is needed on further improving the performance measures. Sentiment analysis can be applied for new applications. Although the techniques and algorithms used for sentiment analysis are advancing fast, however, a lot of problems in this field of study remain unsolved. The main challenging aspects exist in use of other languages, dealing with negation expressions; produce a summary of opinions based on product features/attributes, complexity of sentence/ document, handling of implicit product features, etc. More future research could be dedicated to these challenges.

### **ACKNOWLEDGEMENT**

We would like to thank our Mentor Professor Snehal Kathale for helping and inspiring us throughout the research. We are also grateful to our head of the department, Professor Nileema Pathak for the support. We would also like to Thank Atharva College of Engineering for the resources and facilities provided. Lastly we Thank our families who sustained us with their cheerful and unlimited support (on many levels), not the other way around. Thus — to end on a sentimental note — this work is dedicated to them

## **REFERENCES**

- [1]. Amruta Sankhe and Prachi Gharpure Feature Based Sentiment Analysis for Online Reviews in Car Domain In International Journal of Current Engineering and Technology , Feb 2014
- [2]. Junichi Tatemura .Virtual Reviewers for collaborative exploration of movie reviews. In Proceedings of Intelligent User Interfaces (IUI), pages 272–275, 2000.
- [3]. Loren Terveen, Will Hill,Brian Amento,David McDonald, and Josh Creter. PHOAKS: A system for sharing recommendations. Communications of the Association for Computing Machinery (CACM), 40(3):59–62, 1997.
- [4]. Xin Jin,Ying Li,Teresa Mah, and Jie Tong. Sensitive webpage classification for content advertising In Proceedings of the International Workshop on Data Mining and Audience Intelligence for Advertising, 2007
- [5]. Jeff Zabin and Alex Jefferies. Social media monitoring and analysis: Generating consumer insights from online conversation. Aberdeen Group Benchmark Report, January 2008
- [6]. F. Benamara et al., “Sentiment Analysis: Adverbs and Adjectives Are Better than Adverbs Alone,” Proc. 2007 Int’l Conf. Weblogs and Social Media (Icwsm 07), 2007.
- [7]. C. Cesarano et al., “Oasys: An Opinion Analysis System,” Proc. AAAI 2006 Spring Symp. Computational Approaches to Analyzing Weblogs, AAAI Press, pp. 21–26,2006.
- [8]. Akshi Kumar et al., “Sentiment Analysis on Twitter” IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012 ISSN (Online): 1694-0814
- [9]. Gilad Mishne and NatalieGlance. Predicting Movie sales from blogger sentiment. In AAAI Symposium on Computational Approaches to Analyzing Weblogs (AAAI-CAAW), pages 155–158, 2006.