



ISSN: 0975-833X

RESEARCH ARTICLE

SENTIMENT ANALYSIS – INSIGHT INTO CUSTOMER OPINION

Prof. Dr.P.K. Srimani* and Srinivas, A**

*Former Chairman, Department of Computer Science and Maths, Bangalore University, India

**Faculty of Computer Science, Surana College, Bangalore, Karnataka, India

ARTICLE INFO

Article History:

Received 27th November, 2011
Received in revised form
24th December, 2011
Accepted 18th January, 2011
Published online 29th February, 2012

Key words:

Sentiment analysis,
Opinion mining,
Customers,
Web content mining, Tools.

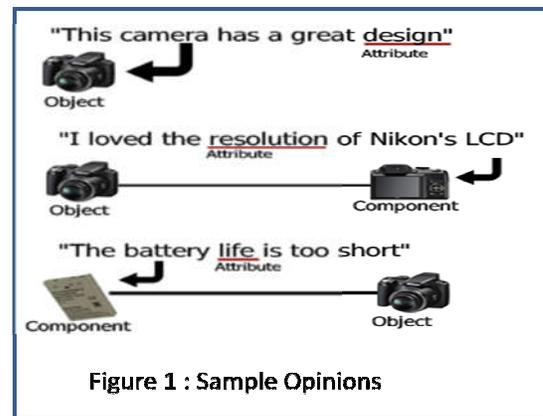
ABSTRACT

This paper an overview of the state of the art in the area of sentiment analysis is presented. Research is being carried out in this area. The approaches and techniques of sentiment analysis and its application in various fields are present here. Within we focus, on automated sentiment analysis using Natural Language Processing (NLP). In addition some sentiment analysis tools available to pinpoint the effect of specific issues on the customer perceptions are discussed, helping the vendors respond with appropriate marketing and PR strategies. It is concluded that the present work provides a platform for further research.

Copy Right, IJCR, 2012, Academic Journals. All rights reserved.

INTRODUCTION

With the explosion of Web 2.0 platforms such as blogs, discussion forums, peer-to-peer networks, and various other types of social media all of which continue to proliferate across the Internet at lightning speed, consumers have at their disposal a soapbox of unprecedented reach and power by which to share their brand experiences and opinions, positive or negative, regarding any product or service. As major companies are increasingly coming to realize, these consumer voices can wield enormous influence in shaping the opinions of other consumers—and, ultimately, their brand loyalties, their purchase decisions, and their own brand advocacy. Companies can respond to the consumer insights they generate through social media monitoring and analysis by modifying their marketing messages, brand positioning, product development, and other activities accordingly. Sentiment analysis also called opinion mining or voice of the customer is a field of Web Content Mining that aims to find valuable information out of users opinions. Mining opinions on the web is a fairly new subject, and its importance has grown significantly mainly due to the fast growth of e-commerce, blogs and forums. From the business perspective getting important information out of opinions can represent a good source of advertisement or product feedback. For example, a web site specialized on electronics reviews could place advertisements on their pages based on consumer opinions. For instance, if the majority of users express negative opinions about a given product, the web site could place from an alternative product from a competitor. Also, manufacturers can get the feedback of their products to improve their products or services.



Automated sentiment analysis is an emerging field that overlaps with many others such as business intelligence, customer service, and brand reputation management, and the market is hard to measure.

- Opinion mining (sentiment mining, opinion/sentiment extraction) is the area of research that attempts to make automatic systems to determine human opinion from text written in natural language.
- It seeks to identify the view point (s) underlying a text span; an example application is classifying a movie review as thumbs up or thumbs down.

Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" often with an implicit category of "neutral". An example is shown in fig. 1.

*Corresponding author: asrhod@gmail.com

Sentiment analysis is the process that “aims to determine the attitude of a speaker or a writer with respect to some topic.” Extensive research is being carried out in this area. Automated sentiment analysis is the process of training a computer to identify sentiment within content through Natural Language Processing (NLP). Various sentiment measurement platforms employ different techniques and statistical methodologies to evaluate sentiment across the web. Some rely 100% on automated sentiment, some employ humans to analyze sentiment, and some use a hybrid system. With the explosion of Web 2.0 platforms such as blogs, discussion forums, peer-to-peer networks, and various other types of social media, consumers are able to share their share their brand experiences and opinions, positive or negative, regarding any product or service. Many types of sentiment software use a technology known as text analytics, which extracts insight from text, such as in social media, news articles, or internal documents and databases. The market for text analytics alone may rise to \$978 million in 2014 from \$499 million in 2011, according to an October 2009 report by Forrester Research (FORR).

APPLICATION AREAS

Voice of the Voter: Sentiment analysis helps political organizations, campaigns, and news analysts better understand which issues and positions matter most to voters. The technology was applied during the 2010 British and American national election campaigns. Expect it to break big later this year as 2012 U.S. presidential and congressional campaigns ramp up.

Capital Markets Modeling : Market sentiment is part of the common vocabulary -- think bullish and bearish -- with attitudes inferred from news, trading patterns, technical indicators, and good-old gut feeling. Traders are on a constant look-out for an edge, for insights and a jump on the market that provide advantage. Automated financial (and consumer) news analysis, with extracted sentiment corrected to market models are some of the techniques being applied by the trading companies.

Voice of the Market: It is about understanding aggregate opinions and trends. It's about knowing what stakeholders -- customers, potential customers, influencers, whoever -- are saying about the companies and their competitors', products and services. Automated sentiment methods complement human judgment to create more accurate and timely pictures of the Voice of the Market, for competitive intelligence and product development and positioning.

Voice of the Customer: It looks at individual customers, at their service and support needs and issues. VOC and Voice of the Market techniques are similar. One needs monitoring and analysis tools -- but sources differ to an extent, as does the application focus. The opinion mining system could be potentially used by casual Internet users. For a mentioned feature level analysis (see Sec. 5) can be a very good way (if accurate) to provide a summarized view of posts for community review sites (for instance movie reviews or product reviews on Amazon book store). Such small enhancements can greatly improve user experience thus being beneficial for both content consumers and producers. VOC done right will draw data from the full set of customer touch points -- from e-mail, surveys, and contact-center notes in

addition to social sources -- and match customer voices to transactions (inquiries, purchases, returns) and individual customer profiles captured in enterprise operational systems.

Voice of the Employee: Text analytics is a huge VOE enabler, just as it is for understanding voices in other types of surveys.

Brand Reputation Management: Sentiment analysis is an essential component of brand reputation management. It involves listening (monitoring, text analytics, presentation) technologies to online and social media. These are places where anyone -- past/current/prospective customer, self-styled know-it-all, industry authority -- can post opinions that can damage or boost the company's reputation. . It reacts to and attempts to shape perceptions rather than to manage experiences.

Online Commerce: The key to selling a product is responding to customers demands in proper time and in the right location. Many companies spend huge money on market analysis and hire external specialized consulting companies. The opinion mining techniques could aid this effort and potentially minimize costs. Market analysis done by specialized companies is needed to take certain amounts of time and effort, while in many cases getting fast access to accurate market data can be a key factor. The right opinion mining tools could create a business advantage for a company to get ahead of its competitors and swiftly react to customer needs. The assumption is that consumers value others' opinions, about travel, restaurants, stores, and products, opinions presented both in narrative form and reduced to aggregate ratings. Opinions guide us in searches, hence Bing- and Google-computed star ratings. Opinions also help us find products and content that appeal to the customer once he reaches out to a site.

SENTIMENT ANALYSIS MODEL



Figure 2 A typical Sentiment **Analysis Model**

Fig 2 shows a typical sentiment analysis model. The different customer reviews are collated and then the data preparation step performs necessary data preprocessing and cleaning on the dataset for the subsequent analysis. Some commonly used preprocessing steps include removing non-textual contents and markup tags (for HTML pages), and removing information about the reviews that are not required for sentiment analysis,

such as review dates and reviewers' names. The review analysis step analyzes the linguistic features of reviews so that interesting information, including opinions and/or product features, can be identified. This step often applies various computational linguistics tasks to reviews first, and then extracts opinions and product features from the processed reviews. Two commonly adopted tasks for review analysis are POS tagging and negation tagging. The next task involves sentiment classification. There are two main techniques for sentiment classification.

- Subjective (opinion) vs. Objective (fact) sentences
- Positive (favorable) vs. Negative (unfavorable)

DOCUMENT LEVEL SENTIMENT ANALYSIS

Document opinion analysis is about classifying the overall sentiments expressed by the authors in the entire document text. The task is to determine whether the document is positive, negative or neutral about a certain object. When applied to a single type of text those techniques typically have a range of accuracy from 70% to 80% depending on amount of human input and type of text [3].

SENTENCE LEVEL SENTIMENT ANALYSIS

The sentence level opinion mining is an action that can be associated with two tasks. Initial work is to identify whether the sentence is subjective (opinionated) or objective. The second task is to classify a subjective sentence and determine if it is positive, negative or neutral. Similarly as with document level most techniques use forms of machine learning.

FEATURE BASED SENTIMENT ANALYSIS

One of the basic tasks in opinion mining or sentiment mining is classifying the polarity of a given text or feature/aspect level to find out whether it is positive, negative or neutral. Different methodologies are used for this purpose. Some expert analysts used the scaling system to associate numbers with appropriate sentiments that a word is depicting. Research has also shown that subjectivity or objectivity identification can also achieve the purpose. However the most fine grained analysis model would be the feature or aspect based sentiment mining method for this purpose. The basic idea of feature based sentiment mining is to determine the sentiments or opinions that are expressed on different features or aspects of entities. When text is classified at document level or sentence level it might not tell what the opinion holder likes or dislikes. If a document is positive on an object it clearly does not mean that the opinion holder will hold positive opinions about all the aspects or features of the object. Similarly if a document is negative it does not mean that the opinion holder will dislike everything about the object described.

Tasks in Feature Based Opinion Mining

When evaluating documents like product reviews mostly the opinion holders write both positive and negative reviews about a product or service. The general sentiment may be evaluated with the mood of the document to be positive or negative. However in order to obtain such detailed aspects in a text feature level evaluation is required.

There are generally three key types of tasks involved in feature based opinion mining:

1. **Identifying the Object Features:** If we are analyzing a review of a product like digital camera and it says that 'the picture quality of this camera is good' the object feature here will be picture quality. This is the technique involved here to identify nouns and noun phrases as features. Here information extraction techniques can be valuable like Conditional Random Fields or Hidden Markov Models.
2. **Determining the Opinion Orientation:** The second task is to determine whether the feature that is outlined as positive, negative or neutral. This can be slightly complicated in some sentences but achievable generally. For example, the opinion on the feature picture quality is positive. Many approaches can be used in this regard. A commonly known approach is the lexicon based approach for determining the orientation of an opinion or feature in sentences. Another approach like relaxation labeling based approach can also be used.
3. **Grouping Synonyms:** After determining the orientation, the third task is to group the synonyms. The object features can be expressed with different words and phrases.

ARCHITECTURE OF OPINION MINING SYSTEM

Most approaches to sentiment analysis (i.e., classification) of documents essentially adhere to more or less similar frameworks consisting of creating a list of words and their associated sentiment from a training corpus and a subsequent method for scoring documents. Sentiment may be scored on document level, sentence level, or window level. In this process, most approaches rely on a wordbank, typically containing per-word sentiment scores. Creation methods include supervised learning on a set of manually rated documents, learning through related word expansion, completely manual creation, or a combination of these methods. Part-of-speech (POS) tagging is considered to be helpful in sentiment analysis, as it can be used to create algorithms to distinguish sentiment-carrying words like adjectives or adverbs. An Opinion mining framework suggested by [6] is shown in the figure 3. The system counts with a crawling module, which first downloads all the reviews and stores them in the database. After that a POS tagger tags all the reviews which will work as hooks for the mining part responsible for finding frequent features.

SENTIMENT ANALYSIS TOOLS

Below listed are few free online tools for simple sentiment analysis to pinpoint the effect of specific issues on your customer perceptions, helping you respond with appropriate marketing and PR strategies.

Sentiment Analysis Tools Based on Twitter Tweets

- Tweet feel scours Twitter for Tweets about brands of your choice and shows you how positively or negatively Twitter users feel about it. It's a quick way to get a feel for what people are thinking about your brand/product.

- Twendz is a Twitter mining web application that utilizes the power of Twitter Search, highlighting conversation themes and sentiment of the tweets that talk about topics you are interested in (such as your brands).
- Twitrratr searches Twitter for a keyword and the results it gets back are cross referenced against its adjective lists, and then displayed accordingly.

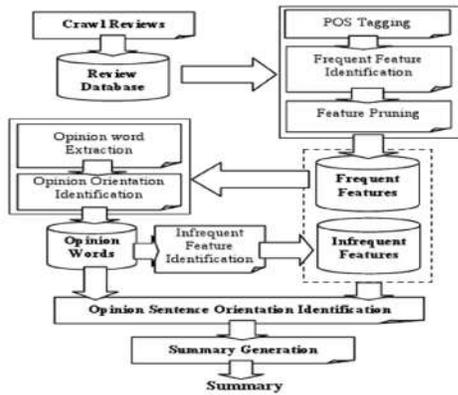


Figure 3: Architecture of Opinion Mining System

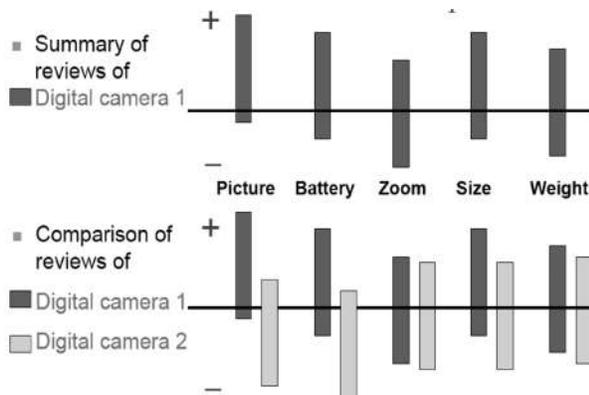


Figure 4: Sample output of the feature based sentiment analysis[3]

SENTIMENT ANALYSIS TOOLS

Below listed are few free online tools for simple sentiment analysis to pinpoint the effect of specific issues on your customer perceptions, helping you respond with appropriate marketing and PR strategies.

Sentiment Analysis Tools Based on Twitter Tweets

- Tweet feel scours Twitter for Tweets about brands of your choice and shows you how positively or negatively Twitter users feel about it. It's a quick way to get a feel for what people are thinking about your brand/product.
- Twendz is a Twitter mining web application that utilizes the power of Twitter Search, highlighting conversation themes and sentiment of the tweets that talk about topics you are interested in (such as your brands).
- Twitrratr searches Twitter for a keyword and the results it gets back are cross referenced against its adjective lists, and then displayed accordingly.

Sentiment Analysis Based on Online News

- **Newssift** (a product from Financial Times) indexes content from major news and business sources and annotates all of this content (and excludes content that lacks credibility or business relevance). It matches business topics to the query (e.g. brands, legal risks, the launch of a new product, environmental impact) so that one can know about changing issues across time for a company or product.

Mood Analysis based on LiveJournal Posts

- **MoodViews** is a collection of tools for tracking the stream of mood-annotated text made available by LiveJournal. At present, MoodViews consists of three components, each offering a different view of global mood levels, the aggregate across all postings of the various moods:
- **Moodspotter** lets you search for mood-associations in a period: you select the period and type in a word or phrase, and Moodspotter spots the top moods associated with this word or phrase in the selected period for you.
- **Moodgrapher** tracks the global mood levels, you can select the mood level (sad, sick, worried,...) with date interval.

CONCLUSION

In this paper we have explored the benefits of sentiment analysis and different implementation models are explored and in particular the feature based sentiment analysis model. Much research has been done to find methods of improving feature extraction to a refined level. The methods of ontology and lexicon have been useful in this regard but still work is being done to improve the overall effectiveness of feature based opinion mining. It is concluded that the present work provides a platform for further research.

REFERENCES

1. Sentiment Classification bibliography, <http://iinwww.ira.uka.de/bibliography/Misc/Sentiment.htm>
2. Opinion Mining bibliography by Jan Wiebe, <http://www.cs.pitt.edu/~wiebe/subjectivityBib.html>
3. Liu, B., Opinion Mining and Summarization, World Wide Web Conference, Beijing, China, 2008
4. Turney, P., "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews.", ACL 2002, 417-424
5. Bo Pang and Lillian Lee and Shivakumar Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques", Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002, 79-86
6. Philip S. Yu Ding Xiaowen, Liu Bing. A holistic lexicon-based approach to opinion mining. WSDM'08, 2008. vii, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 39, 58, 59, 87.
7. Berger, A., Della Pietra, S., J. Della Pietra, V., "A maximum entropy approach to natural language processing", Computational Linguistics, 1999
