MARKETING RESEARCH: THE ROLE OF SENTIMENT ANALYSIS

1. SENTIMENT ANALYSIS: METHODOLOGY

A graphical description of the processes involve in sentiment analysis is detailed in Figure 1

Figure 1: Sentiment analysis process

1.1 Data collection

Sentiment analysis takes advantage of the vast user generated content over the internet. The data source points to queries of user discussions on public forums like blogs, discussion boards and product reviews boards as well as on private logs through social network sites like Twitter and Facebook. Very often, the data log is bulky, disorganized, and disintegrated on multiple portals.

Opinions and feelings are expressed in different ways including the amount of details given, type of vocabulary used, context of writing, slangs and lingua variations are just a few examples. This makes manual analysis tedious, and almost impossible. But, with sentiment analysis, innovative text analytics and natural language processing is employed to extract and classify data. Once the data is extracted, it will then be prepared for analysis. An example of the data is presented in Figure 2.

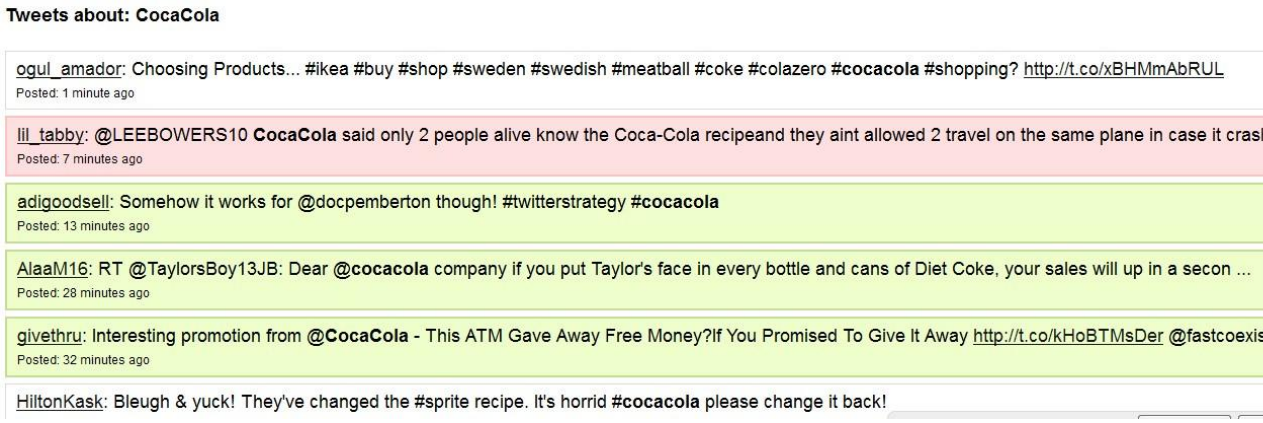

Figure 2: Data extracted from online sources
Source: http://www.sentiment140.com/search?query=CocaCola&hl=en

1.2 Text preparation

Text preparation involves cleaning the extracted data before the analysis is performed. Usually text preparation involves identifying and eliminating non textual content from the textual dataset, and any information that can reveal the identities of reviewers including: reviewer name, reviewer location, review date. In addition, any other content that is not deemed relevant to the area of study is also removed from the textual dataset such as includes stop words or words that are not relevant to the course of analysis.

Presentation of output

The general purpose of the analysis is to convert unstructured fragmented text into meaningful information. Once the analysis is completed, a number of conventional options are used to display the result of text analysis. Chief among them is the use of graphical displays such as pie charts, bar charts and line graphs. The polarity is segmented on color, frequencies, percentages and size. The format of presentation depends on the research interest. Examples of each are presented in Figure 3-7 below:

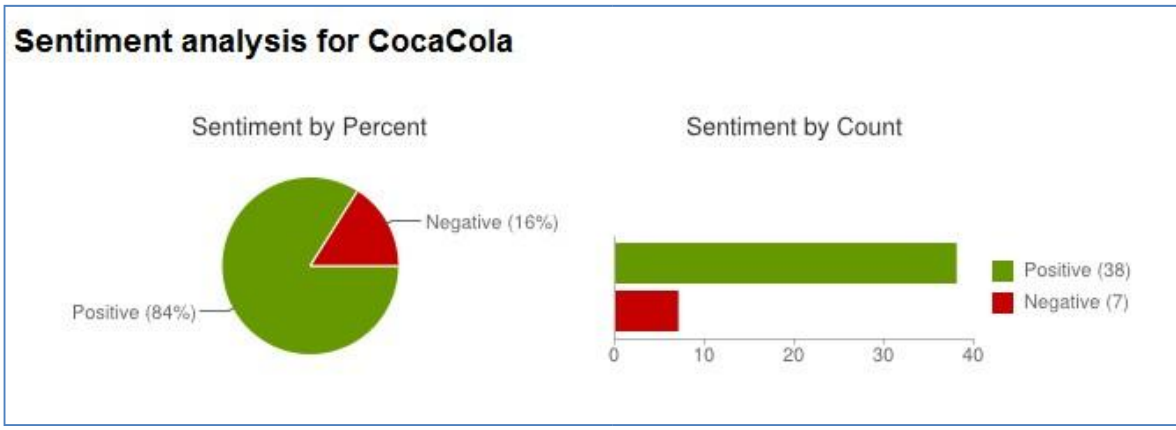

Figure 3: Summary of sentiments on a single product Source: http://www.sentiment140.com/

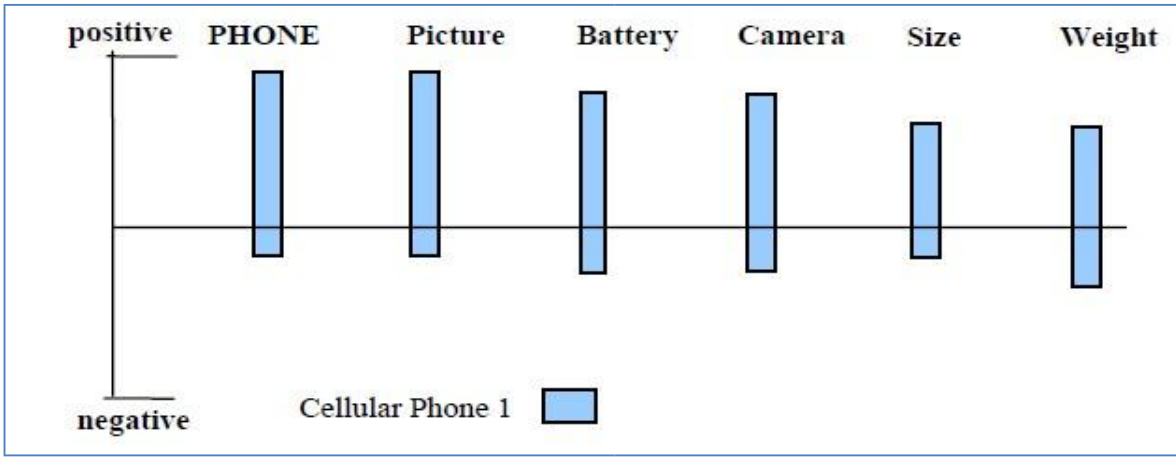

Figure 4: Summary of sentiments on features of a single product (Source: Lu, 2010 p.8)

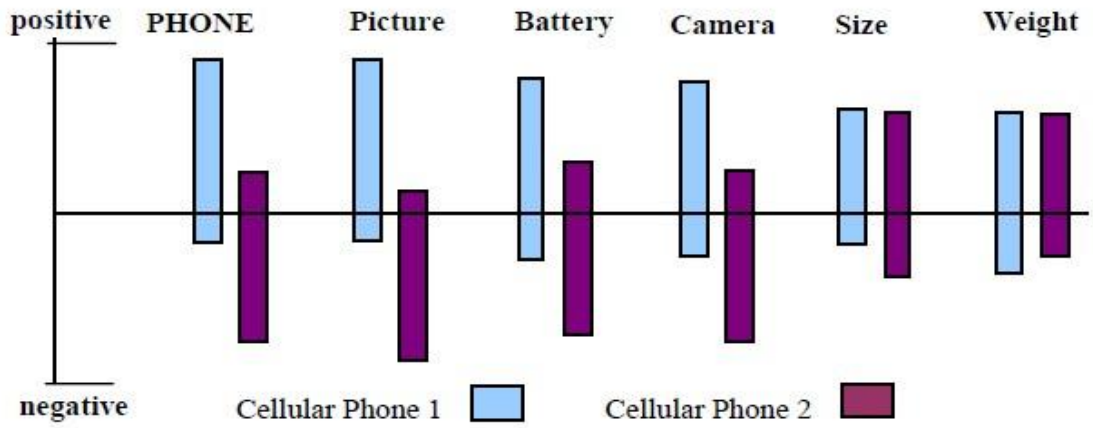

Figure 5: Summary of sentiments comparing two products (Source: Lu, 2010 p.8)

Ratings are also presented by the number of stars with the number of reviews next to it. This is shown in Figure 6.

Figure 6: Summary of sentiments comparing using stars and ratings (Source: http://www. amazon.com)

Time can be included in the analysis. Usually, this is graphically displayed through constructing a sentiment time line by plotting the value of the chosen

statistic (example frequency, percentages, and averages) over time. This is shown in Figure 7.

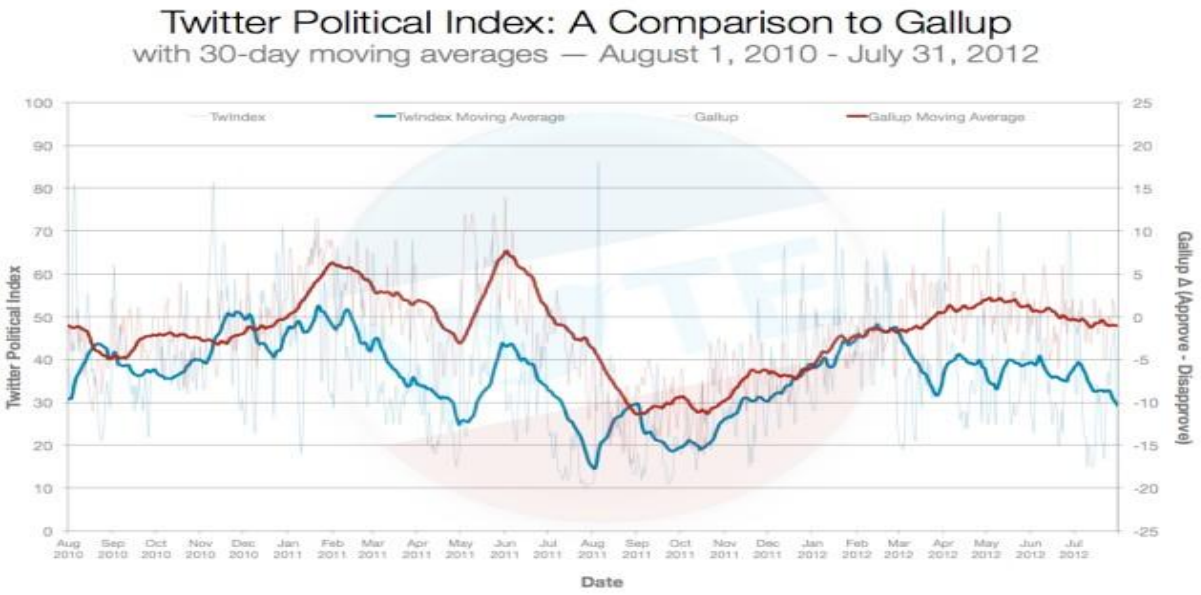

Figure 7: Summary of sentiments over time (Source: http://techcrunch.com/2012/08/01/twitter-launches-its-own-political-barometer-totrack-u-s-presidential-elections/)

2. USES OF SENTIMENT ANALYSIS

Relative to uses, most published work on sentiment analysis focus on product analysis, i.e. extracting opinions on a specific product like phones, movies, or hotels. However, this interest is being extended to include more abstract areas like product features and attributes i.e "feature extraction". Li (2005) defines features as "components or attributes of a specific object". Feature extraction in sentiment analysis is a process that dissects a product into several features or attributes which are used as topic sentences for extraction and classification. This method gives a detailed analysis of sentiments highlighting aspects of a product a customer are pleased with and aspects they are disenchanted with. It also gives meaningful insights into why customers feel the way they do. Additionally, work is being conducted on analyzing statements from multiple targets. This is called comparative sentences and rank objects based on preference.

Although the concept is relatively new, the use of sentiment analysis in a commercial environment is growing. This is evident in the increasing number of brand tracking and marketing companies offering this service. Some services include:

- Tracking users and non-users opinions and ratings on products and services
- Monitoring issues confronting the company so as to prevent viral effects
- Assessing market buzz, competitor activity and customer trends, fads and fashion

- Measuring public response to an activity or company related issue

3. TOOLS AND WORKS IN SENTIMENT ANALYSIS

According to Pang and Lee (2008), researchers have found ways to avoid the use of manual annotation by utilizing existing online textual content generated from sites such as Epinion, Amazon, Rotten Tomatoes, Twitter, Facebook. Several sentiment search engines exist where users run typical queries on any topic of interest, and generate text results. Usually the results are coded and categorized into two or three polar categories. Some examples currently available are:

1. Twitrratr – www.twitrratr.com
2. Sentiment 140 - http://www.sentiment140.com
3. Tweetfeel – www.tweetfeel.com
4. Opinmind – www.opinmind.com
5. Social Mention – www.socialmention.com

Sentiment search engines make sentiment analysis quite easy. But, the online reviews on sites like Amazon and Epinion have been found to be skewed towards the positive which raises questions on validity and reliability of sentiment classification. However, Pang and Lee (2008) admit that although the content might be skewed, the validity of the process is acclaimed. Another tool in sentiment analysis is word lists or annotated databases which categorize words based on their emotions for example -attractive (positive valance) or aversive (negative valance). Some examples include: ANEW, General Inquirer and LIWC. Other tools include sentiment analysis programs that are specifically designed to categorize short textural documents. One example is ***sentistrength***.