

Measuring brand sentiment using smileys

Sharat Dwibhasi, MS in MIS, SAS and OSU Data Mining Certificate,

Dr. Goutam Chakraborty, Oklahoma State University, Dr. Johnson Joseph, University of Miami

Abstract:

It is of paramount importance for brand managers to measure and understand consumer brand associations and the mind-space their brand captures. Brands are encoded in memory on a cognitive and emotional basis. Traditionally, brand tracking has been done by surveys and taking feedbacks, resulting in a direct-communication which only covers cognitive segment and misses emotional segment. Respondents generally behave differently under observation and in solitude.

In this project, a new brand tracking technique is proposed which involves capturing public data from social media that thereby focuses more on emotional aspects. For conceptualizing and testing this approach, we downloaded around one million tweets for three major brands – Nike, Adidas and Reebok posted by the users. We proposed a methodology and calculated metrics (benefits and attributes) using this data for each brand.

We noticed that generally emoticons are not used in sentiment mining. To incorporate them, a macro was created that automatically cleans the tweets and replaces emoticons with an equivalent text. We then built supervised and unsupervised models. The results depict that using emoticons improves efficiency of predicting the polarity of sentiments as misclassification rate decreased from 0.31 to 0.24. Using this methodology, we tried to track the reactions that trigger in the minds of customers, when they think about a brand and thus analyze their mindshare.

I. Introduction:

The million dollar secret for any business to succeed is to understand their customers, who in a broader sense are none other than– Humans! Based on this mantra, marketers and researchers as early as 1900's came up with studies that relate human psychology behind the inclination towards buying a product. In short, that was how branding was evolved. Today although with the evolution of Internet and advancement of technology, everything changed including the customers, the products, and the way they buy but one thing has remained fairly constant – the psychological needs of the customers.

To understand whether a brand is actually satisfying the psychological needs of the customers, millions of dollars are spent each year all around the world. It is of paramount importance to brand managers to measure and understand consumer brand associations and the responses that are evoked when they think about the brand.

As Keller quotes in his book, “The power of a brand lies in what resides in the minds of the customer” [1], it is evident that brand perception and associations are cognitive constructs placed inside consumer memory. Traditionally consumer behavior is cast in terms of cognitive psychology. It implies that it is often looked from a perspective that people/consumers involve in elaborate, conscious information processing before they think/decide to buy a brand [2]. Well, this holds true for majority of the cases but research proves that unconscious emotions impact the buying patterns towards a brand. Unfortunately the concept of emotion in marketing literature is quite nebulous making marketers to overlook emotions that are fundamentally believed to be unconscious.

Brands are encoded in memory on a cognitive (thinking, analytical, considered) and emotional (somatic) basis. These two elements of brand encoding are intricately linked and it is emotional coding rather than reasoned argument or perceived quality that determines whether or not people take notice of the stimuli related to the brand, such as direct communications [3]. This brings us to the point that, when brand tracking is done using the traditional approach of carrying out surveys and taking feedback then marketers are involving in a direct-communication which only covers cognitive segment.

In this light, we try to propose a brand tracking methodology that involves both these cognitive and emotional segments.

II. Methodology:

The cognitive information can be retrieved from the usual techniques of surveys or interviews, which is not new to marketers and researchers. In addition to these, in this research paper the main focus would be on extracting useful information from the social media and utilizing it to build a brand association network that resides inside the minds of consumers.

With the advancement of technology, people are spending more and more time on social media websites thereby generating massive public data by sharing their views and opinions on everything under the sun (such as issues/products/brands) over various platforms such as social media, bulletin boards, blogs and product reviews. The reactions (in the form of like/dislike/comment/share) to things they see over the internet is unconsciously guided by the associative network residing in their minds.

Numerous studies all over the world that have been published lately, highlights how this online data can be leveraged to get valuable information. We will start extracting data from one of those social media giants – twitter and later the scope of this study could be expanded to include data from other available platforms as well. This study proposes a process (as shown in Figure: 1) for brand tracking which includes two major phases. The phase 1 (Brand Image) constitutes of Benefits and Attributes.

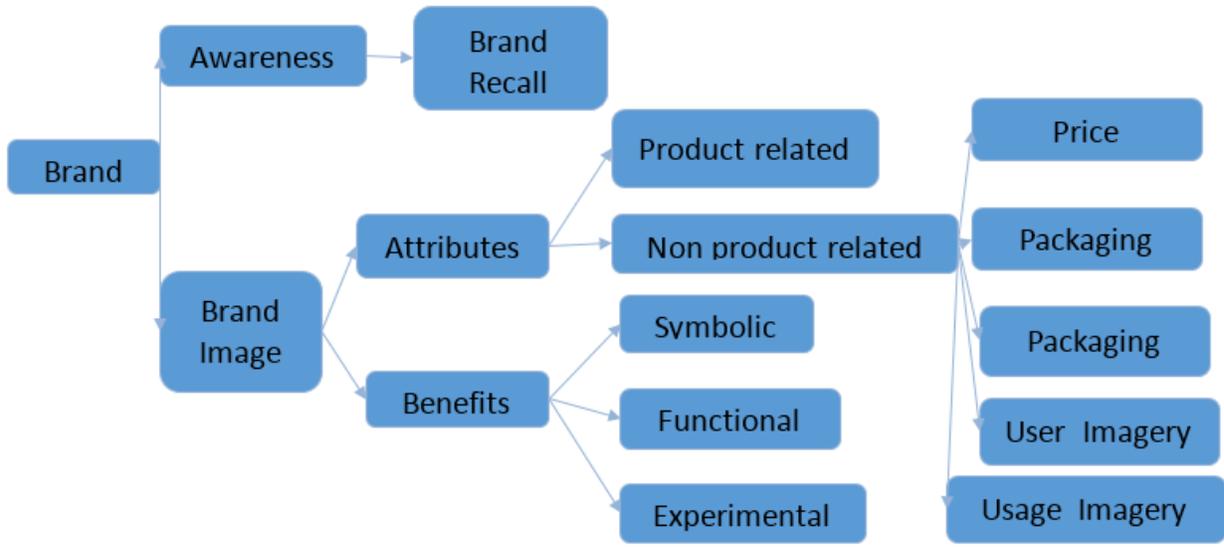


Figure 1: Brand Tracking Tree

For the purpose of our study, we downloaded real-time data (tweets) available over the internet for three major brands namely – Nike, Adidas and Reebok. We downloaded tweets starting from May 2015 for these three brands and extracted information that would help us to map the brand. Starting with the lowest limb of the tree from the brand Image node, we need to calculate a) words associated with the brands “Nike”, “Adidas” and “Reebok” b) Classify these words into *attributes* and *benefits* based on the parts of speech.



Figure 2: Tweets displaying functional benefit

The users in Figure 2 tweet about how comfortable their shoes are and as “Comfort” is a noun, it can be classified as a functional benefit. The frame-work for generating this is illustrated in Figure 3 and it includes 1) Extracting product dimensions (like performance) from text that users may not explicitly mention but inadvertently talk about 2) determining the overall sentiment of the user 3) Using these variables to determine/track brand performance.

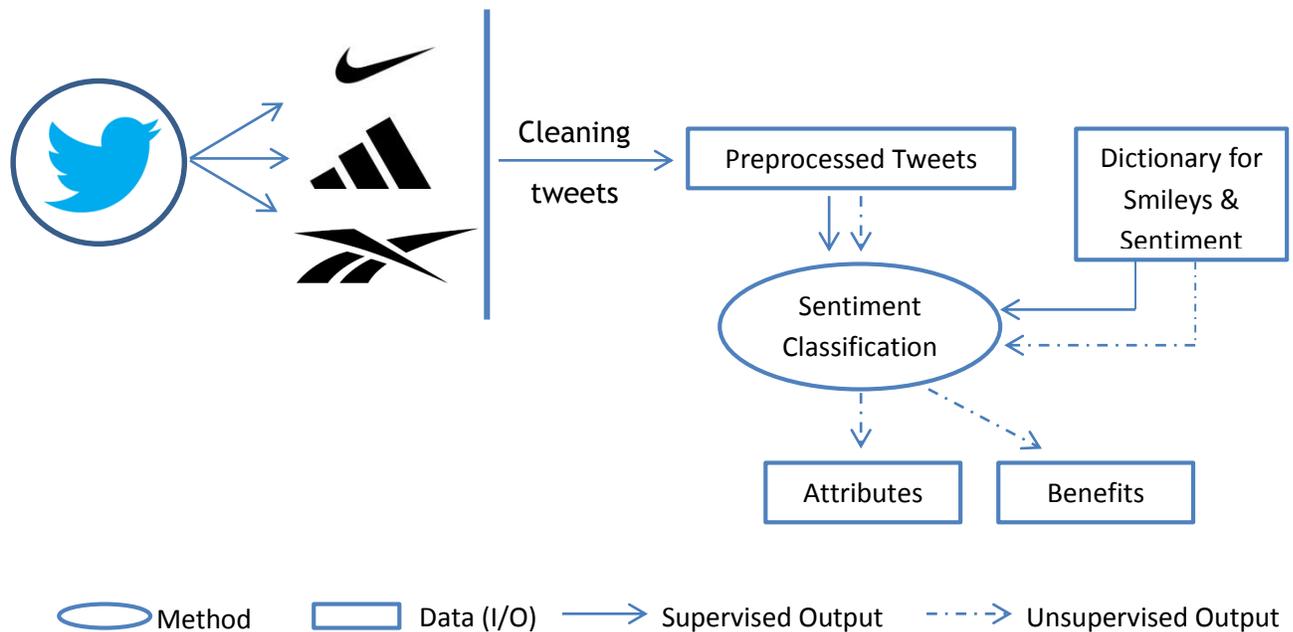


Figure 3: Methodology that was followed

III. Data Cleaning:

As for any data analysis project understanding the data is the most important step, we spent first three months of the project in cleaning and understanding the data itself. We first tried to collect data from Facebook and Twitter but owing to the recent changes in Facebook, we are unable to pull data from any open page/group as now that data access is restricted to administrators of the page. But fortunately, twitter API allows downloading/tracking of tweets for a given hashtag.

We used a crawler (using Google Script) to extract and save tweets for a given hashtag over the cloud. The crawler extracted information provided by the Twitter Streaming API and the variables we got in our dataset are as shown in the table below:

| Variable | Description |
|--------------|--|
| Date | Date and time of the tweet posted |
| Twitter User | User Id of twitter User |
| Followers | Number of followers for the user |
| Follows | Number of people that user follows |
| Retweets | Number of retweets for the tweet user posted |
| Favorites | Number of people who favorited the tweet |

| | |
|--------------|--|
| Tweet Text | 140 character text tweet |
| Lang | Language in which tweet was posted |
| Source | Device used to post this tweet |
| Geo | Geographic location of the tweet |
| Location | Coordinates of the user |
| Description | Status text of the user |
| Created At | Date and time of the user account creation |
| Status Count | Number of tweets posted |

Table 1: Metadata of tweets

Now, after thorough inspection and discussion, we listed down the issues in the data that needed our attention. The points of concern that were listed out are as mentioned below:

- Every tweet that was retweeted contained the pre-fix RT and we had to remove this.
- Almost every tweet contains a link, which from the text-mining perspective is not required and hence can be deleted
- Most of the tweets have a username (@username), if they are tagging someone and it becomes important to remove them as a part of text cleaning.

We used some of the pearl functions and SAS® to create a macro that automatically cleans tweets (assuming the datasets are in the same format as the Twitter API). This macro saved us a lot of manual efforts and labor. The macro programming is attached as an appendix to this research paper.

IV. Extracting information from Emoticons:

In the textual world, there are two types of Emoticons:

- Textual Emoticons - :)
- Unicode Emoticons - ☺

Nowadays, most of the applications and websites have the feasibility of automatically converting textual to Unicode emoticons (emoji's/smiley's). It was a point of concern mainly because of two reasons:

- It was easier to decipher the textual emoticons and replace them with equivalent text
- Text mining tools like SAS cannot handle Unicode Emoticons and they simply ignore them in a text.

We started by converting the textual emoticons by replacing them with an equivalent text but realized that around 90% of the smileys used were Unicode emoticons and to predict the sentiment accurately emoticons can play a major role. Interpreting emoticons can help us to predict the sentiment of tweets more accurately. There were many tweets which seemed to be neutral when they were considered without emoticons but when we added information from the emoticon, it changed the sentiment completely.



Figure 4: Tweet without the Emoticon

Considering the tweet in above figure 4, there is nothing substantial that could lead us to predict the sentiment of this user accurately.

Whereas, if we consider this tweet along with the emoticon as shown in figure 5, we can definitely claim that he is in love with the brand

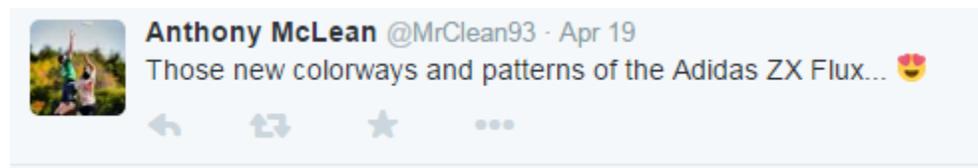


Figure 5: Using emoticons from the tweet makes interpretation accurate

Typically, text-mining tools fail to interpret smileys and we often lose valuable information which is embedded in the form of smileys. Thus understanding and extracting information from the Emoticon becomes the need of the hour. As a solution, we created a dictionary of more than 1000 Emoticons with their definitions and have replaced these emoticons with their corresponding meaning using Python. The emoticon dictionary and the Python code for replacing emoticons are attached as an appendix to this research paper.

V. Results:

We have got the results from the first phase where-in based on the words we will categorize them into attributes and benefits. The words and their corresponding roles, inverse-document frequency, term frequency and number of docs are as shown in the Table 2 below:

Table 2: Top 10 words along with their metrics for each brand

| Adidas | | | | | Nike | | | | | Reebok | | | | |
|-----------|------|------------|----------------|-------|--------|------|----------|----------------|-------|------------|------|----------|----------------|-------|
| Term | Role | IDF | Term Frequency | #Docs | Term | Role | IDF | Term Frequency | #Docs | Term | Role | IDF | Term Frequency | #Docs |
| adidas | Prop | 1.57974994 | 52266 | 50741 | nike | Prop | 1.290321 | 77855 | 74693 | reebok | Prop | 1.269734 | 39973 | 37454 |
| adidas | Noun | 2.47458237 | 30621 | 27289 | air | Noun | 2.959186 | 23727 | 23491 | deal | Noun | 4.014708 | 5611 | 5587 |
| size | Noun | 3.59682481 | 12596 | 12536 | size | Noun | 3.405451 | 17395 | 17241 | shoe | Noun | 4.098705 | 7529 | 5271 |
| nba | Prop | 4.27563344 | 10237 | 7831 | nike | Noun | 3.732597 | 14093 | 13743 | jersey | Noun | 4.113838 | 5793 | 5216 |
| shoe | Noun | 4.55195648 | 7199 | 6466 | shoe | Noun | 3.908558 | 12583 | 12165 | size | Noun | 4.250374 | 4785 | 4745 |
| jersey | Noun | 4.5998201 | 6363 | 6255 | jordan | Prop | 3.990288 | 11676 | 11495 | reebok | Noun | 4.384415 | 4375 | 4324 |
| deal | Noun | 4.86770624 | 5235 | 5195 | run | Verb | 4.10166 | 11340 | 10641 | bestseller | Noun | 4.568506 | 3806 | 3806 |
| boost | Noun | 5.0469646 | 4966 | 4588 | roshe | Prop | 4.940818 | 5975 | 5948 | auction | Noun | 4.604585 | 3712 | 3712 |
| black | Noun | 5.08261995 | 4698 | 4476 | black | Noun | 4.942516 | 6038 | 5941 | nfl | Prop | 4.832754 | 3969 | 3169 |
| superstar | Noun | 5.32035 | 3890 | 3796 | max | Noun | 4.966019 | 5932 | 5845 | classic | Adj | 4.847396 | 3140 | 3137 |

For calculating the favorability, we need to measure the polarity of the sentiments. We built two rule based models – one considering the emoticons and one without the emoticons. Results show that the model utilizing emoticons yields better efficiency.

| Target | Statistics Label | Training | Validation |
|------------------|------------------------|----------|------------|
| Sentiment | Misclassification rate | 0.27 | 0.36 |
| Sentiment Smiley | Misclassification rate | 0.24 | 0.30 |

VI. Further Work

The scope of this project can be easily extended to gather and interpret data from other social media websites.

With the guidance of Dr. Goutam Chakraborty and Dr. Joseph Johnson, I am working on completing this project and calculate the remaining nodes for the brand tree map.

VII. References:

- [1] Keller, K. L. (2008). Strategic Brand Management - Building, Measuring and Managing Brand Equity (3rd Edition ed.). New Jersey: Pearson Education International.
- [2] Dijksterhuis, A., Smith, P. K., van Baaren, R. B., & Wigboldus, D. H. (2005). The Unconscious Consume: Effects of The Environment on Consumer Behaviour. Journal of Consumer Psychology , 193-202.
- [3] Gordon, W. (2001). The darkroom of the mind - What does neuropsychology now tell us about brands? Journal of Consumer Behaviour , 280-292.

VIII. Contact Information

Your comments and questions are valued and encouraged. Contact the author at:

Sharat Dwibhasi, Oklahoma State University, Stillwater OK

E-mail: Sharat.dwibhasi@okstate.edu

Sharat Dwibhasi is graduate student majoring in Management Science and Information Systems at Oklahoma State University. He has been using SAS® tools over two years for Data Mining, Texting Mining, and Sentiment Analysis projects. He has published research papers in SAS Global Forum 2015, SAS Analytics Conference 2015 and SCSUG, 2014 Proceedings. He has also won the SAS Student Scholarship award in 2015 and was among the top eight teams selected in the SAS Student Symposium 2016.

Dr. Goutam Chakraborty, Oklahoma State University, Stillwater OK

Email: goutam.chakraborty@okstate.edu

Dr. Goutam Chakraborty is Ralph A. and Peggy A. Brenneman professor of marketing and founder of SAS® and OSU data mining certificate and SAS® and OSU business analytics certificate at Oklahoma State University. He has published in many journals such as Journal of Interactive Marketing, Journal of Advertising Research, Journal of Advertising, Journal of Business Research, etc. He has chaired the national conference for direct marketing educators for 2004 and 2005 and co-chaired M2007 data mining conference. He has over 25 years of experience in using SAS® for data analysis. He is also a Business Knowledge Series instructor for SAS®.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.