



<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *Conference on Data Science and Knowledge Engineering for Sensing Decision Support (FLINS 2018), Belfast, Northern Ireland, UK, 21 – 24, August 2018..*

Citation for the original published paper:

Giri, C., Thomassey, S., Zeng, X. (2018)

Analysis of consumer emotions about fashion brands: An exploratory study

In: Jun Liu (Ulster University, UK), Jie Lu (University of Technology Sydney, Australia), Yang Xu (Southwest Jiaotong University, China), Luis Martinez (University of Jaén, Spain) and Etienne E Kerre (University of Ghent, Belgium) (ed.), *Proceedings of the 13th International FLINS Conference (FLINS 2018): World Scientific Proceedings Series on Computer Engineering and Information Science* (pp. 1567-1574).

World Scientific Proceedings Series on Computer Engineering and Information Science

https://doi.org/10.1142/9789813273238_0195

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:hb:diva-22821>

ANALYSIS OF CONSUMER EMOTIONS ABOUT FASHION BRANDS: AN EXPLORATORY STUDY

CHANDADEVI GIRI ^[1, 2, 3], NITIN HARALE ^[1]
SEBASTIEN THOMASSEY ^[1], XIANYI ZENG ^[1]

¹ Gemtex, Ensait, 2 Allée Louise et Victor Champier, 59056 Roubaix, France

² University of Borås, SE-501 90 Borås, Sweden

³ Soochow, College of Textile and Clothing Engineering, Suzhou 21506, China

Fashion products are characterized by high variability in terms of rapidly changing consumer preferences. Consumers express their emotions on social networks such as Twitter, Facebook and Instagram. The main objective of this paper is to explore Twitter data for recognizing customer sentiments about fashion brands and to analyze their overall perception towards the brands. Two brands, Zara and Levis, are considered and users' tweets related to these brands are analyzed using text mining and Naïve Bayes classifier. The results from this study suggest that social media such as Twitter can serve to be the repository of consumer sentiments and opinions. Sentiment analysis of the tweets can indicate fashion trend and thereby enable fashion brand companies to quickly respond to the ever changing consumer demands.

Key words: Sentiment analysis, fashion industry, big data, Twitter

1. Introduction

Fashion industry is one of those industries which are striving to exploit “Big Data” technology for the real time prediction of customer demands, their choices and sentiments about the products. Due to the advent of e-commerce, digital businesses, and the social media, consumers are rapidly voicing their preferences and opinions about the fashion products all over the internet than ever before. Due to ever-changing consumer preferences and demands, it is indispensable for the fashion brand companies to analyze the market trend in a real time. However, it poses significant challenges to the fashion designers and manufacturers to track the consumer choices and market trend as they are increasingly leaving their footprints on social media such as Twitter, Facebook, Instagram, etc. Therefore, fashion industry is flocking to “Big Data” analytics and web semantics tools to accurately forecast demands of their customers, and to prevent them from switching to other brands. Many researchers in the domain of fashion technology have argued for the extensive analysis of social media

platforms to make the flawless forecasts using effective “Data Analytics” tools to spot the real time fashion trends [1]. It is in this context, this paper aims to extract the consumer sentiments from their tweets on Twitter platform.

Twitter is an online social networking and blogging service that was founded in 2006. One of the major applications of Twitter analysis has been the prediction of political election results in major countries such as USA, UK, Canada, Australia, etc. [2]. Unlike the election results, fashion industry does not depend on the fixed time market trend prediction given the rapid fluctuations in the consumer demands and choices, and therefore it is a great challenge to extract real time information from the Twitter data.

To address this challenge, this paper is an attempt to detect the real time perceptions of customers about fashion brands by analyzing Twitter data. We randomly chose only two well-known fashion brands, Zara and Levis, for the sake of small scale study. However, more number of brands could be studied for the same purpose. We applied Naïve Bayes classification method to classify emotions from the tweets as it requires only a small amount of trained data. For the computational part, we used an open-source statistical software, RStudio because of its flexible functionality for exploratory work.

The rest of the paper is structured as follows: In the second section, we present a brief overview of current approaches in the social media data analysis for fashion industry. In the section 3, we present the theoretical research framework, in which research questions, data collection and processing is described. Experimental results are discussed in the section 4, and the final section 5 presents the future scope of the study.

2. Related Literature

Ortigosa et al [3] suggest that the companies should create their marketing campaigns based on the understanding of consumer’s emotional attachments toward all the aspects of the brands and thereby provide them customized products. Vishal Vyas et al [4] tested the performance of Rapid Miner and machine learning algorithms such as SVM, Naïve Bayes, Decision Trees by using them to derive the consumer sentiments from the tweets. Besides other prominent industries, fashion industry is investing in analytical tools and technology infrastructure required for social media analytics. However, sentiment analysis is one of the topics that remain to be the part of mainstream Big Data analytics and it warrants a further research in order for it to help today’s digital businesses to survive the fierce competition. Consumer sentiment analysis using social media data could become the key part of both short and

long term strategic planning of the fashion industry and it could enable them to retain and expand their customer base and make them happy and satisfied [5].

In line with the above-mentioned researches, we argue that sentiment analysis could help fashion industry to understand their customers' demands, their grievances, and product preferences in a real time to be able to gain a competitive advantage in a today's digital business era.

3. Research Framework

It is important for fashion industry to be innovative to strike the right spot for managing its brand value. Nowadays, Consumers are more brand conscious than before and their choices are changing quickly. Any outfits get outdated within a month. Thus, it is essential for the fashion industry to integrate their business with the advanced big data tools to get the real time valuable insights and exploit the opportunity of this Internet era.

As online users are increasing dramatically, our aim is to derive sentiments from the tweets of consumers focusing on fashion brands and to classify emotions and polarity of tweets.

To reach this aim, we address the following research questions;

1. Are the emotions in the tweets useful to know the popularity of fashion brands?
2. Can we identify the public opinion in the tweets about the brands?

To address our research question we choose two brands for our study 'Zara' and 'Levis'. Figure 1 depicts the research framework for this paper. The steps are followed by tweets extraction, cleaning tweets and finally doing sentiment analysis on two fashion brands.

3.1. Data Extraction and Preprocessing

Data is extracted from Twitter API using direct authentication with the help of "ROAuth" package in RStudio. We followed the standard method of fetching Twitter data by creating an app via which we could request API to Twitter. However, Twitter observes and strictly restricts users from accessing high load

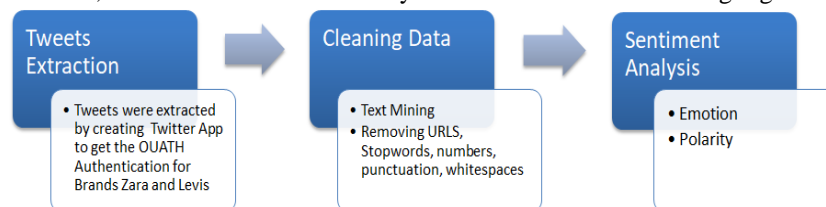


Figure 1: Research Framework

API, owing to which, we could fetch Tweets for only last 10 days. While fetching tweets, we used hashtags to get precise information about the brand Zara and Levis for 10 days (from 18-02-2018 to 28-02-2018). During these 10 days, we collected 702 and 980 tweets for #Zara and #Levis respectively. As we know that the tweets data are messy and it is quite difficult to work with such cluttered data, therefore, data cleaning is the first step in text mining. Text mining was applied to remove hashtags, URLs, punctuation, retweets and whitespaces from the tweets data. As a result, the no. of tweets of the two brands is reduced to 534 and 586. Figure 2 illustrates how tweets are transformed after cleaning.

Tweets Before Cleaning
[1] "Personal shopper summer superstars from #Zara Every now and then zara excels, right now there are https://t.co/ud1x0vMOz1 #fashionblog"
[2] "Exclusive look at Zara, Raceway Mall, New Jersey, united states https://t.co/NjqG5IoJLV via @YouTube @ShopFreehold... https://t.co/tLgus3NcNt "
Tweets after Cleaning
[1] "personal shopper summer superstars from every now and then zara excels right now there are"
[2] "exclusive look at zara raceway mall new jersey united states"

Figure 2: Tweets data before and after cleaning

3.2. Sentiment Analysis for Brand Zara and Levis:

Machine learning algorithms such as Naïve Bayes, SVM (Support Vector Machines), k-NN are popular techniques for opinion mining and text categorization from Text data [6]. However, we chose Naïve Bayes algorithm because it is advantageous over other methods since it requires a small amount of training data for parameter estimation. Naïve Bayes is a probabilistic classifier and it works on the principle of "Bayes' Conditional Probability Theorem." Bayes' rule is presented in Eq. 1.

$$P(Y|X) = \frac{P(Y)P(X)}{P(Y)} \quad (1)$$

We have a dataset $T = \{t_1, t_2, t_3, \dots, t_n\}$, which includes n tweets. For each given tweet in Twitter data "T", Naïve Bayes calculates posterior probability of classes $c \in C$, and assigns the class, \hat{C} , having the maximum posterior probability to the given tweet.

$$\hat{C} = \operatorname{argmax} P(c|t) = \operatorname{argmax} \frac{P(t|c)P(c)}{P(t)}, \text{ where } c \in C \quad (2)$$

It classifies the tweets based on the score that is calculated for the emotions present in the tweets. We trained Naïve Bayes on the dataset that is created by [8], and it is comprised of approximately 1542 words and class labels of six emotion categories such as “joy”, “sadness”, “anger”, “surprise”, “fear,” and “disgust.” The score of a word “w” in a tweet “t” for a class “c” is represented as in Eq. 3.

$$\text{score}[c]=\log \text{ prior } [c]+\log \text{ likelihood } [w,c] \tag{3}$$

Where, $\log \text{ prior } [c]=\log \frac{N_c}{N}$ and $\log \text{ likelihood } [w,c]=\log \frac{\text{count}(w,c)+1}{\sum \text{count}(w,c)+1}$
 N= total counts of words in Dataset and N_c = counts of words from Dataset in each class c [7]. And, for each word w in L (L = Lexicon of dataset D), $\text{count}(w,c)$ = number of occurrences of w in Tweets and $\text{count}(w',c)$ = number of occurrences of w in L.

Emotion class having a maximum score for a given tweet out of six emotions is the best fit which can be seen in the below figure 3. If no word in a tweet matches with the dataset [8], the best fit is assigned as “NA” which we defined as ‘unknown’ category.

▲ ANGER	DISGUST	FEAR	JOY	SADNESS	SURPRISE	BEST FIT
5 1.46871776464786	3.09234031207392	2.06783599555953	1.02547755260094	7.34083555412328	2.78695866252273	sadness
6 7.34083555412328	3.09234031207392	2.06783599555953	1.02547755260094	1.7277074477352	2.78695866252273	anger

Figure 3. Best-fit emotion based on maximum score

Tweets devoid of any of these emotions are then categorized as “unknown.” Emotions were classified and clustered according to categories of emotions for the brands Zara and Levis and it can be visualized in the Figure 4 and 5.

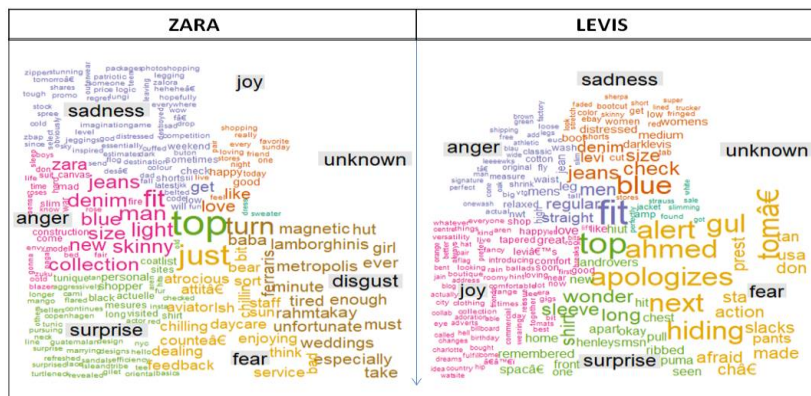


Figure 4. Word Cloud for emotions for the brand Zara and Levis

As we can see in Figure 5 that there are significantly many tweets in the “unknown” category. Approximately 75% and 61 % of the tweets for the brand Zara and Levis respectively fall under “unknown” category. It could be attributed to the fact that the dataset we used is relatively small and the trained data entail only 1542 words. With the help of wordcloud clustering, as shown in Figure 4, we can see that words in the tweets are expressing customer sentiments for the products of the two brands. For example, as seen in Figure 4, emotion “anger” expresses certain features of brand Zara, while for Levis brand it expresses a few different features.

The categories of emotions as depicted in Figure 5 can further be classified as an extension in order to discern the overall polarity of emotions such as “Positive” or “Negative” or “Neutral.” as is shown in Figure 7. We used Janyce Wiebe’s subjectivity lexicon [8], for the classification of polarity in the tweets. We calculated log likelihood of each tweet considering that it either belongs to “Positive,” “Negative” or “Neutral” class. The value of the ratio “likelihood of

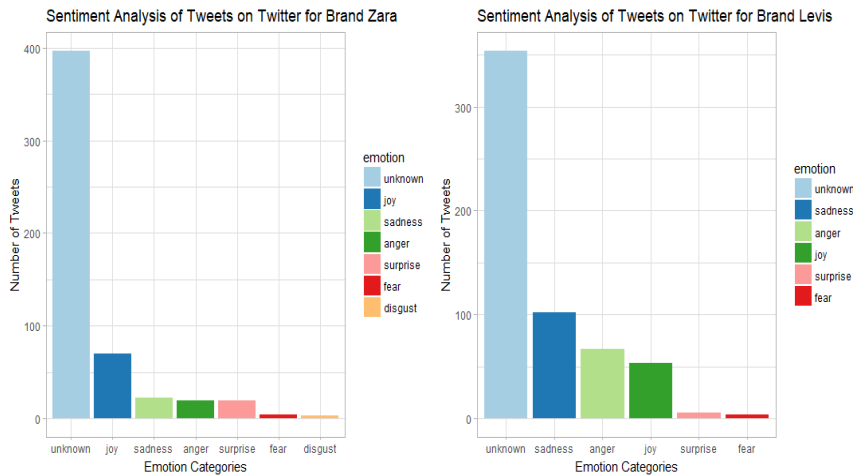


Figure 5. Sentiment Analysis of Tweets for Brand Zara and Levis

Positive/likelihood of Negative” determines the class for each sentence. For example, for the brand Levis, the best fit for a polarity can be calculated as below.

if score > 1 = positive, if score < 1 = negative; if ratio score = 1 then its neutral

	POS	NEG	POS/NEG	BEST_FIT
5	17.9196623384892	17.1191924966825	1.04675862146897	neutral
6	17.2265151579293	0.445453222112551	38.6718836070664	positive

Figure 6: Best-fit polarity score for a tweet

These categories could help designers and marketing analysts to have the clear picture of the polarity of their consumers' opinions. As we can see from the Figure 6, both brands have "Positive," "Negative," and "Neutral" tweets and their popularity is evident from the high number of positive tweets in 10 days period.

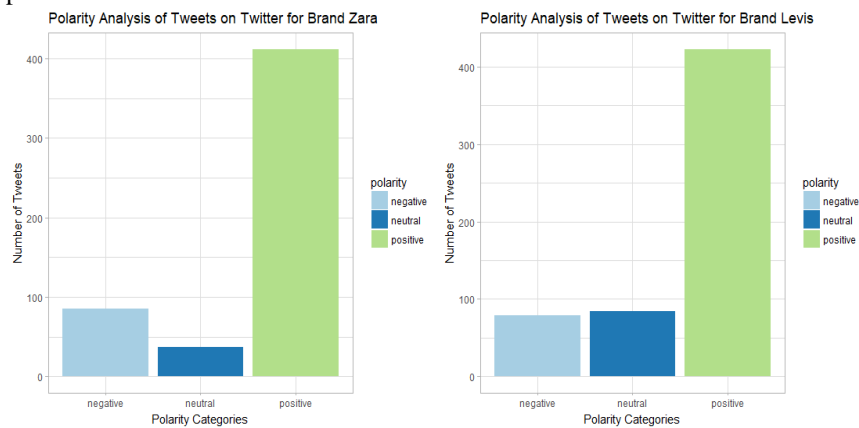


Figure 7: Polarity Analysis of Tweets on Twitter for Brand Zara and Levis

4. Results, Discussion and Conclusion

We extracted the tweets for the brand Zara and Levis. After we applied data cleaning and text mining on tweets, the number of the tweets of Levis and Zara was reduced significantly. Sentiment analysis was performed on tweets for emotions and polarity for both the brands. The tweet emotions were clustered into seven categories, and we then analyzed the words related to each category as shown in the figure 4. It is important to note that the algorithm was unable to classify the emotions for approximately 75% tweets and 61% tweets of the brands Zara and Levis respectively. Such tweets are categorized as "unknown". It could be attributed to the problematic trained data in the sense that words included in the trained data are limited and it is not representative of all kinds of emotions. Moreover, the words devoid of expressions or the words used for certain product names, for ex. "Jeans," "Clothes", would not possibly be classified as emotions. This paper is not aimed at comparative analysis of the popularity of two brands; however, by identifying the polarity of emotions in the tweets towards "Positive," "Negative" and "Neutral," popularity of the brands could be monitored. Furthermore, the polarity classification indicates that the users are quite positive about both the brands because more than 70% tweets fall under positive category. In conclusion, sentiment analysis could be performed

on the tweets and it is possible to identify consumer opinions about the fashion brands. Positive emotions could be the indicator of popularity of the brands. Sentiment analysis is the part of big data analytics and investment in such tools could enable fashion industry to improve design, production and the overall performance of their businesses.

5. Future work

Given the complexity of sentiment analysis from social media data and the dearth of advanced research work in this area, we consider this paper to be the initial step towards the advanced research in future. The limitation of this study lies in the fact that tweets for only ten days were used for the analysis and the language of the tweets was English. Moreover, we extracted data using #Zara and #Levis and the result was for overall opinion about the brand and not about specific products. Extracting data could be further improved by using keywords specific to products associated with brands. From Big Data perspective, this work could be extended for seasonal sentiment trend analysis for famous brands if we use historical tweets spanning over a long period of time and also include different regions in the world.

Acknowledgment

We extend our sincere gratitude to the SMDTex – Sustainable Design and Management of Textiles, Erasmus Mundus Joint Doctoral commission for providing us with excellent research environment and consistent support.

References

1. Beheshti-Kashi S, Thoben K-D. *Dynamics in Logistics Lecture Notes in Logistics* (2015).
2. Conway BA, Kenski K, Wang D. J. *Computer-Mediated Communication 20* (2015).
3. Ortigosa A, Martín JM, Carro RM. *Computers in Human Behavior* (2014).
4. Vyas V, Uma V. *Procedia Computer Science* (2018).
5. Wright S, Calof JL. *European Journal of Marketing* (2006).
6. Rezwani M, Ali A, Rahman A. *International Journal of Advanced Computer Science and Applications* (2017).
7. Strapparava C and Valitutti A. *Proceedings of the Fourth International Conference on Language Resources and Evaluation* (2004).
8. Riloff E, Wiebe J. *Proceedings of the 2003 conference on Empirical methods in natural language processing* (2003).