Predicting the consumption behavior media essay

Media

Abstract -It is impossible to deny the blooming popularity of social networking websites because these sites provide a great degree of user intercommunication, social networking and content sharing. These free as well as easy to use cloud based applications provide users and companies a constant stream of valuable communication. With all the hype that surrounds social networking sites this area seems to be perfect platform for extensive data mining, development and research. The content that is generated from the social media used by millions of people on daily basis, if tapped can be used to predict real world outcomes. There is a fair amount of useful information that if recovered and interpreted intelligently can lead to a lot of constructive results. We show that how a simple model based on the rate of creation of positive and negative tweets about smart phones can be utilized to forecast the consumption behavior of smart phones. Keywords: Data Mining, Sales Forecasting, Smart Phones, Social Media, Tweets, Twitter.

I. INTRODUCTION

Social networking websites like Twitter; Face book, and LinkedIn MySpace etc, have evolved as such a category of online discourse that it can be interpreted as a form of collective wisdom. The ease of use, reach and speed offered by these websites has changed the public discourse in society with in no time. They have also set the agendas and trends in topics ranging from technology and fashion to environment, sports and movies. To Gain insight on consumer behavior and attitudes has always been a critical task for huge marketers like Ford Procter & Gamble, Samsung, and Apple etc: but now social media data can be investigated to predict real world outcomes. This technique of predicting consumer attitude is a lot more accurate than

surveys and opinion polls. The social networking websites present a picture of the future by extracting what is exciting and popular on the internet. Since there is enormity and high variance of information along with great level of disclosure routinely engaged in by the websites users; hence these sites act as a fertile ground for harnessing the data into the form that can act as a predictor for certain outcomes. A model can also be built to comprehend the beliefs of the collective population to gain valuable insights into the behavior and to predict future trends. This paper aims to report on such a research. Specifically we focus on the task of examining the distributions of tweets for different smart phones using the chatter from Twitter and finding the correlation with the consumption behavior of smart phones. Twitter chatter is considered to be one of the fastest growing social networks on the internet. Twitter[1]is a micro blogging network having a huge user base and it experiences a burst of popularity among several millions of users actively participating in the formation and propagation of content. The study is focused on smart phones because of two main reasonsThe users of social media find considerable interest in this topic, characterized by large numbers of users discussing about smart phones along with significant diversity in their opinions. The revenue generated by a smart phone can easily indicate the real world outcome. The paper targets at:-Firstly at assessing how buzz and attention about different smart phones is created and how the attention changes with time. We also focus on the fact that how hype created on twitter can forecast real world performance. The hypothesis is that smart phones that are well talked will generate more revenue. We also study that how do positive and negative opinions propagate and how far they can influence general public. If the negative opinions are far above the

positive opinions then people get discouraged to buy a particular Smartphone while some positive reviews can generate curiosity and interest among people.

II. LITERATURE REVIEW

Twitter is one of the most popular web services but it has not been used to comprehend the consumer attitudes and behavior. Francesco [3] and others have used a business process classification framework to lay the research topics in the context of business applications and has provided an indication of key problems and techniques in the area of social network analysis and mining. Java et al [1] performed investigation on community structure and performed grouping on different categories of user intentions on Twitter. Pedro Domingos [4] have build social networking models that allow designing of " viral marketing" plans in order to maximize positive word of mouth among customers. Some prior work has been done on analysis of correlation between blogs and review mentions and performance. Stephen Patton [6] has discussed data mining and investigative techniques for social networking websites. Sitaram and others [7] have used twitter to predict the future with social media. Gruhl and others [2] have shown the process of generating automated queries for mining blogs. Although the research has been there on predicting the sales of smart phones based on meta-data information on the phones such as hardware specifications, powerful CPU, graphics processor etc: Sharda and others [5] have treated the problem of prediction as a classification problem and has used neural networks for classification of movies. We have shown how we can use our model to forecast the consumption attitude of smart phones buyers and how can we

correlate the revenue generated from smart phones with the sentiments generated on twitter.

III. ABOUT TWITTER

Twitter was launched on July 13; 2006 by Jack Dorsey. It is a directed network for real time information that connects you to the latest news, ideas, opinions and stories. It has an extremely large user base that consists of millions of users. At the heart of the twitter reside small bursts of information known as tweets. Tweets are short messages wherein each message is 140 characters long. Each user submits status periodically in form of tweets. These updates consist of real time information such as opinions, news, personal information, images, videos and articles etc: The updates made by a user gets displayed on the profile page of the user as well as his/her followers. It also facilitates sending a direct message to another user. A post that is originally made by one user and that is forwarded by another user is known as retweet. These retweets acts as a tremendous means of propagating interesting links and posts via Twitter communityTwitter aids in connecting businesses to customers in real time. It has an immense potential for viral marketing since due to its huge reach businesses can speedily distribute information to people having interest in their products and services. Twitter or similar micro-blogging services can be used by number of businesses and organizations for new understanding, for dissemination of information to stake holders, for useful predictions, for advertising products, for decision making and for designing viral marketing plans.

IV. DATASET CHARACTERISTICS

The dataset under observation is obtained by crawling twitter. com on hourly basis. The feeds are searched using the keywords based on names of the smart phones ensuring tweets are extracted to the maximum accuracy. Using Tweeter Search API[2], timestamp, name of the author and tweet itself was extracted for analysis. We extracted tweets for five smart phones of different brand to cover maximum market. The tweets were monitored from three months prior to the release to two week after release of phone. We collected the data 3 months prior to release to measure the anticipation and buzz created by it. The data wasn’t restricted to a single country, though country wise tweets were segregated but overall impact was measured. Most of the smart phone brands have a pattern of release. The proclamation followed by final release. Though release of the phone varies from country to country but actual buzz is created only after initial statement. The period one week prior to release and two weeks after release is taken as critical period for measurement. The details of the smart phones selected with their release time are listed in the Table 1. Majority sample is of year 2012, we have also considered phones from year 2011 for broader perspective. We have taken care of the ambiguities by choosing appropriate keyword combinations, also the data was prepared for analysis by filtering. Fig. 1. shows the three month prelease tweets of the dataset. The tweets are categorized as two types influential or non influential. Any tweet which makes an impact on other authors is termed as influential.[8] The tweet with maximum following and retweeting lies in this category. We can observe Iphone 5 had maximum prerelease tweets to his credit and so is the pre-booking pattern [9]. Samsung doesn’t follow the usual strategy of publicizing product before

release but its tweets make it evident its popularity and thus revenue generated by its pre-booking. The pre-booking sales of the dataset shows similar trend as of the tweets. Fig. 2. Shows the 7 days tweet after the official release of the phone. Though the phones were released at different times but the patterns of tweets are similar in nature. The mere fact that apple was unable to produce enough Iphone 5 to cater to all the orders is apparent from the number of tweets recorded after its release. Samsung tweets were consistent and so were its tweets. Though Nokia Lumia 800 pre release attention was not so high but post release the popularity augmented and so did its sales. The phones with lesser impact on authors failed[i]generate the desired revenue [10]. Through tweets we also calculated momentum and acceleration of tweets. Momentum is defined and measured as weighted count of tweets of the keywords. Acceleration of a phone is measured as how rapidly their mention increases or decreases in propagation. Table 2. Compiles the momentum and acceleration of the dataset.

V. CHATTER AND SALES

Though chatter cannot solely depict the sales of products of different price ranges, but it surely shares the direct relationship with the sales [11]. As soon Iphone 5 was released after much created hype by Apple, the tweets were maximum and so were its sales comparable to any other smart phone. Owing to latest law suit between the Apple and Samsung, tweets of both the product were risen. The first spike in weekly growth coincides with the same week that a jury gave its verdict in favor of Apple for infringement of its patent by Samsung. The inundation of post-litigation press coverage and the

deluge of chatters both drove general attention to Samsung and suggested that Samsung devices were similar enough to IPhones to be chosen as its replacements.[3]Our research showed the same behavior in the pattern of tweets. The second spike started the week of September 12, when Apple announced the new IPhone 5. Apple’s announcement was quickly followed by hundreds of tweets comparing both the phones, most of which drew attention to their. Eventually the growth started to slowdown, resulting in 0 % growth suggesting that many consumers were waiting to hear about the new IPhone before making a decision [12]. One of the interesting fact is also that Samsung Galaxy S3 tweets were consistent and cumulative and so were the sales. The average week-over-week growth of 9% means that every week 9% more Galaxy S3 handsets were added than the week before. This also means that the growth effects are cumulative, so while the Samsung device’s weekly new device growth slowed some weeks, it is still adding handsets at a remarkable rate.

VI. OPINION MINING

Opinion mining or sentiment analysis plays an important role in investigating and predicting future outcomes. Opinion mining is used to infer subjectivity from the fact. It classifies an opinion into three categories namely positive, negative and an implicit category neutral. It is a well-studied problem in machine learning, computational linguistic and text analytics. We used an online tool Topsy[4]to classify the tweets also an open source GATE[5]was to extract and analyze the sentiments. We used the trained classifier to predict the sentiments of the tweets and thus calculate subjectivity and polarity of the dataset.

A. SUBJECTIVITY

An objective sentence expresses some factual information about the world, while a subjective sentence expresses some personal feelings or beliefs [13]. Measuring subjectivity can predict whether the user will be keen on buying a product. Consumer buying behavior depends upon subjectivity, more than factual implications. First we classified the tweets as opinionated or not opinionated which simply means subjectivity classification. After which we further classified the opinionated tweets as positive or negative which is called sentence level sentiment classification. Past research has shown that adjectives and adverbs are good indicators of subjectivity and opinions. However, although an isolated adjective may indicate subjectivity, there may be an insufficient context to determine its opinion orientation [14]. Thus we used algorithm which not only worked with adjectives and adverbs but also context word. We captured the subjectivity using the following measure: Here neutral tweets indicate the absence of opinion or subjectivity. Fig. 4. is comparing subjectivity of galaxy S3 and Iphone 5. The subjectivity here is referring to the attachment of authors to a particular phone which further is reflected in the sales also.

B. POLARITY

To compute the sentiments for a smart phone, we measured the ratio of positive tweets to negative tweets. Higher the ratio more is the success of the phone.

VII CONCLUSION

This paper shows that how massive amount of information accessible from blogs and knowledge sharing sites can be used to predict the consumer

behavior and attitude. By crawling the tweets on hourly basis we have constructed a model for predicting the revenues from smart phones in advance of their release. Our model also shows that there lies a strong correlation between the amount of attention given to a smart phone in terms of its tweets and the revenue generated by it in future. We also did opinion mining to analyze the sentiments present in tweets in form of positive tweets, negative tweets and neutral tweets and also demonstrated their efficacy at improving prediction. This work also shows that how substantial information present in the social media expresses a collective wisdom which when tapped properly can predict the future outcomes accurately and can also be helpful in decision making [15].