Monitoring and Analyzing Customer Feedback Through Social Media Platforms for Identifying and Remedying Customer Problems

Sumit Bhatia*, Jingxuan Li†, Wei Peng*, and Tong Sun*
*Xerox Research Centre, Webster, NY
†Florida International University, USA
Email: Sumit.Bhatia@xerox.com, jli003@fiu.edu, Wei.Peng@xerox.com, Tong.Sun@xerox.com

***Abstract*—The tremendous growth and popularity of social media platforms like Twitter, Facebook, etc. provides business organizations an opportunity to monitor the feedback from its customers, identify their problems and take corrective measures. In this paper, we describe a system to automatically monitor and analyze customer feedback through various social media platforms like Facebook, Twitter, etc. and detect issues faced by the customers. Business organizations can use this system to engage with their customers and help alleviate the problems faced by them. The system uses statistical event detection techniques for identifying various customer issues. The system offers a batch version as well as real time version of event detection algorithm depending upon the client's requirements. We also describe a few case studies illustrating the utility of our proposed system for business organizations in identifying issues faced by their customers through social media channels.**

I. Introduction

The emergence of social media platforms like Twitter, Facebook, etc., has provided internet users an effective medium to express and communicate their thoughts with hundreds and thousands of other people. Oftentimes, people use these social media platforms to share their experiences with different consumer products and services used by them and thus, can influence the opinion of other people in their social network about these products and services. Such consumer-to-consumer communication facilitated by social media platforms can greatly impact a company's reputation and sales [5]. A survey of 320 active internet users carried by Barnes [1] revealed that almost three quarters (74%) of users choose companies/products based on other customers' experiences shared online. Thus, it is crucial for business organizations to consider the social media phenomenon while developing their marketing and promotional campaigns [8] and customer care strategies.

Various social media platforms like Facebook, Twitter etc., also offer business organizations a convenient means to gather customer feedback about the products/services offered by the company as well as the company's perception among the consumers. For this purpose, many organizations have official Facebook pages or Twitter accounts where consumers can interact with the organizations' representatives (customer care agents), can post their problems and provide details about their experience with the products/services offered by the company. Gathering customer feedback through social media is superior to traditional feedback gathering techniques (e.g. customer satisfaction surveys) owing to its real-time nature and the coverage of a diverse user population. Moreover, social media provides much greater volume of user feedback in a cost efficient way as compared to traditional methods. For example, consider a mobile service provider that wants to gather feedback from its users located in different geographical locations and belonging to different age groups. Performing a user survey or gathering user feedback through telephone/postal mail or even e-mail will first require identifying users in target populations, contacting them and conducting the survey through phone/mail/e-mail etc. This whole exercise will require investment of considerable labor and capital from the company. However, the costs to the company can be greatly reduced if there is an automated tool available to the company that can provide the required information by monitoring the customer feedback in different social media platforms.

In addition to providing useful feedback to the company, the customer comments or posts on social media channels can also be used to identify major problems different consumers might be facing. The company can then take appropriate corrective measures to solve customers' problems. For example, suppose a lot of customers are facing issues related to network coverage for a mobile operator and post their experiences on a social media platform (e.g. Twitter). If the mobile operator can automatically monitor such user comments in real-time and can work with the concerned consumer to alleviate their problems. In this work, we describe *our system that can help business organizations automatically monitor customer feedback through various social media channels* like Twitter, Facebook etc. Further, the system analyzes the feedback gathered to *detect potential problems the customers may be facing*. Upon detection of such problems, the customer care agents can engage with the customers to alleviate their respective problems. We propose to detect the consumer complaints/problems by framing the problem as an *event detection* problem. The system continuously monitors the customer feedback coming through various social media channels and analyzes them in real time to discover *events of interest* to the business clients. We use statistical event and anomaly detection methods to identify

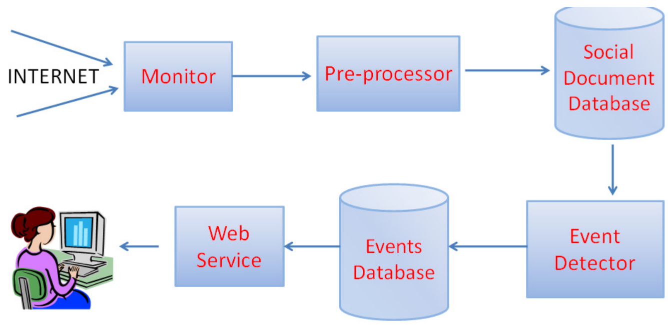


Fig. 1: Major components of the proposed framework.

such events.

II. Proposed Framework

The basic architecture of our proposed system to automatically monitor customer feedback through a social media platform is illustrated in Figure 1. The different components of the system are discussed in more detail below.

1. Monitor: The "monitor" consists of a crawler to crawl publicly available posts made on a social media channel(Twitter, Facebook etc.) that are relevant for a business client. The crawler regularly crawls the social media website and gathers data relevant to the business client based on criteria specified by the client. The relevant documents could be the posts made on the organization's official Twitter handle/Facebook page. Further, the client can also specify a set of keyword related to its products/services and the crawler will then crawl the posts containing these keywords to create an initial set of social media documents related to the client. For example, a mobile operator can specify keywords such as *network, iphone, android, wireless, 4g,* etc. and the crawler can then crawl public tweets mentioning these keywords and the brand name.

2. Pre-processor: The pre-processor performs linguistic processing on the crawled data for further analysis. We use GATE toolkit [2] to identify POS (part of speech) tags in the crawled documents. Further, we perform sentiment analysis of the crawled documents to identify whether the document talks about the concerned product in a negative, positive or neutral tone. For sentiment analysis, we use machine learning based sentiment classifiers developed in-house. The processed documents along with their detected sentiments are stored in the database.

3. Event Detector: This module analyzes the processed social documents to detect events in real time (details about event detection algorithm are in the next section). The detected events and anomalies are then stored in an event database.

4. Web Service Front-end: The customer care agents for the the business client can access the events database through a web service. They are provided a list of important events as detected by the system along with details about the users affected through the event. The customer care agents can then engage with the respective customers to help resolve the problems faced by them.

III. Event Detection

The dictionary[1] meaning of event is, *"something that happens; an occurrence, especially one of some importance"*. The problem of *Event detection* refers to the task of automatically detecting events from the input data. Event detection has various practical applications such as surveillance, scientific discovery, fault detection, anomaly detection, etc. [4]. In our particular problem, the input data consists of the text of social documents being monitored by the system and the "occurrences" correspond to the terms and phrases in the social documents. Out of these occurrences, the important words/phrases constitute "occurrences of importance" or *events*. Thus, the problem of event detection reduces to identifying terms/phrases of interest to the end-user (customer care agent). Thus, in our proposed framework, a detected event corresponds to a term/phrase as observed from the monitored documents.

Driven by practical considerations identified during our discussions with business clients , our framework has two different options for event detection – *(i)* the **batch version** which analyzes a given set of documents at one time, and, *(ii)* the **streaming** or **real-time** version which continuously processes the incoming social documents and detects events in real time. Before going into details of these algorithms, we discuss the following important concepts that lead to the development of said algorithms.

A. Sentiment Analysis of Social Documents

As discussed before, the purpose of the proposed system is to detect events in real time from social documents and bring them to the attention of business clients so that they can engage with their customers and help solve the issues faced by them in a timely manner. Based on our interactions with the business clients and their customer care agents, it was decided that not all events are of equal importance to customer care agents. From a customer care agent's perspective, the events that are of most importance are the ones appearing in the posts with negative sentiments. A post with negative sentiment indicates negative emotional states (such as angry, sad, etc.) of the post author (consumer). On the other hand, the customer care agent may not care that much for the post with positive sentiment. A post with positive sentiment denotes a customer with positive feelings about the company, e.g., happy, satisfied, etc.

B. Relative Importance of Different Terms in Social Documents

Given the large user base and the huge volume of social media data, it is possible that the event detection algorithm detects a large number of events from the input data. However,

[1] http://dictionary.reference.com/browse/event

not all the events are equally important and some events might be more interesting to the client and some may not be that interesting. Thus, we need a metric to measure the relative importance of different events. A traditional approach to define the importance of a term in a document set is TF-IDF (term frequency - inverse document frequency), which is usually defined as follows.

$$\begin{aligned} & TFIDF(term, doc, Docs) \\ & = \sum_{doc \in Docs} tf(term, doc) \times idf(term, Docs), \end{aligned} \tag{1}$$

where tf denotes term frequency, while idf indicates inverse document frequency.

This definition measures the relative importance of a term in a document in terms of its frequency in every document. In our case, each document also has an associated time-stamp indicating the instant of time stamp when this document was posted on the web sites. Further, since the major goal of our event detection framework is to capture current issued faced by users, we argue that in addition to the number of occurrences of a term, the time of occurrence of a term is also an important measure of term importance in our application scenario. In order to account for the time factor, we propose a time normalized version of the TF-IDF metric as defined below.

Hence, the time decay information is adopted into the definition of TF-IDF as:

$$\begin{aligned} & TFIDF(term, doc, Docs) \\ & = \sum_{doc \in Docs} (tf(term, doc) \cdot decay) \times idf(term, Docs) \\ & = \sum_{doc \in Docs} (tf(term, doc) \times \mathrm{e}^{-(currentTime - documentTime)}) \\ & \quad \times idf(term, Docs) \end{aligned} \tag{2}$$

The decay function as described above is chosen so as to assign a higher weight to recent documents. In the following sections, whenever we mention TF-IDF, it is TF-IDF in Equation 2.

Further, as discussed previously, given the practical considerations of treating documents with positive and negative sentiments separately, we chose to compute two separate TF-IDF scores for terms by using the "positive"/"negative" TF and IDF for each term/phrase. The positive (negative) term frequency of a term/phrase is defined as the number of times this term/phrase appears in a document with positive (negative) sentiment. Likewise, the positive (negative) document frequency of a term/phrase indicates the number of documents with positive (negative) sentiments containing this term/phrase. Thus, for each term/phrase, we can have both positive and negative TF-IDF scores, illustrating their importance in documents with respective sentiment labels.

C. Event Detection in Our Framework – New and Anomalous Events

The events as identified by our framework can further be categorized as **New** or **Anomalous** events. A "new" event corresponds to those events that have not been seen before previously and have large number of recent occurrences in documents crawled. For customer care agents, such new events are crucial as they correspond to new issues being faced by the consumers for which a solution does not exist in the customer care knowledge base. Hence, these issues require special consideration from agents. As an example, let us again consider the mobile service company and suppose we found an event "signal", which was not previously observed in the historical data collected for the company. Such a discovery indicates that sufficient number of customers are facing issues getting proper signal and the company can thus take appropriate corrective measures in a timely fashion. We also require a minimum frequency threshold for a term/phrase for it to be considered as a new event. This serves two purposes –*(i)* it prevents noisy terms (spelling mistakes etc.) being identified as new events, and, *(ii)* ensures that the detected new events at least have a sufficient volume.

An "anomalous" event on the other hand refers to those events which may have been seen before but experience a sudden burst in a specific time period. From a customer care agent's perspective, these events represent a sudden increase in the occurrence of a known issue. In order to perform anomalous event detection, both batch and streaming versions of our framework rely on the Grubbs' test [9] which was originally proposed for outlier detection. For a specific event represented by a noun/phrase, given the negative term frequencies as observed during different time periods, Grubbs' test can help us find out the "outlier" time period where the event's frequency was statistically higher than its historical average. We then say that this term is an anomalous events in the given period owing to its anomalous behavior as compared to its historical behavior.

D. Batch version of Event Detection Process

Sometimes, business customers are interested in identifying events in the dataset at hand, that may have been collected from various sources. For such requirements, it is required to process all the documents within the dataset at once and produce the results for the customer. The batch processing procedure as implemented in our proposed system is described in Algorithm 1.

E. Streaming or Real-Time Detection of Events

In many cases, the business customer would like to continuously detect events from the user feedback that is being continuously gathered from different social media websites. Algorithm 2 describes the real-time event detection procedure as implemented in our framework.

To further illustrate the algorithm, Figure 2 describes the complete workflow of the algorithm. The process starts by computing different statistics about terms/phrases and their frequencies from the historical data. The system then continuosly monitors new user feedback at regular intervals. At each new interval, the newly gathered data is analyzed and new and anomalous events are identified. The historical statistics are updated at each interval by incorporating the information gained from analysis of data gathered in this interval.

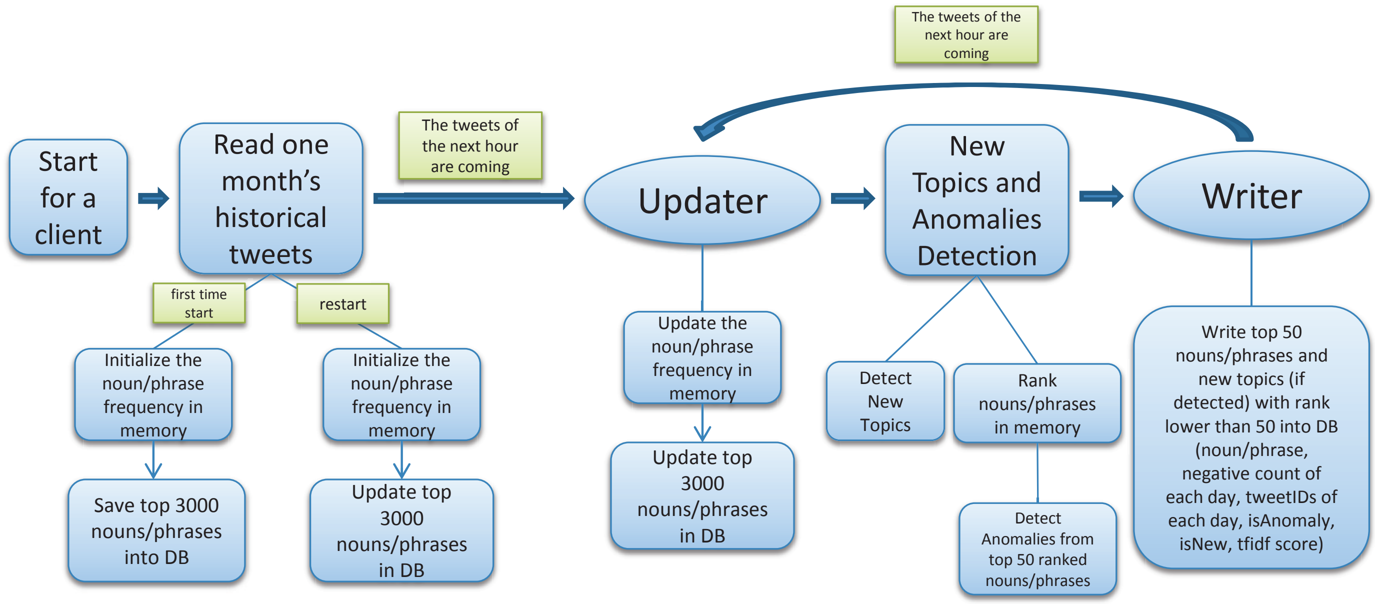


Fig. 2: An example of the streaming algorithm.

Algorithm 1 Batch version

Input: Given D – the set of social documents with their corresponding sentiment labels, t – the time window for which anomalous events are to be detected (1 day in our implementation, events are detected on a daily basis).

1. Perform the sentence splitting, tokenizing, stemming, POS tagging for each sentence in documents in D.

2. Filter the preprocessed tweets by removing stop words, and dirty words.

3. Extract all nouns and noun phrases.

4. Compute time normalized TF-IDF score for each extracted noun phrase and rank them by their time normalized TF-IDF scores.

5. Obtain the negative term frequency within every possible time window, t (according to the time range indicated by the first and the last tweets of the dataset) for each noun phrase.

6. Detect the anomalous noun phrases using Grubbs' test based on the data from step 5.

Algorithm 2 Streaming version

Given a set of historical tweets with the corresponding sentiment, t – the time window for detecting events, T – the historical window, ut – the time window for updating the historical nouns/phrases, hn – the number of top ranked nouns/phrases from the historical data

Read tweets within the historical window T into memory.
if FirstTimeStart **then**
  Initialize the top hn nouns/phrases
else
  Update the top hn nouns/phrases
end if
while NewTweetsComing **do**
  Read tweets within the latest ut into memory.
  Update the top hn based on the new tweets in the memory.
  Detect new events
  Detect anomalous events
end while

IV. Case Studies

In this section, we present a few illustrative case studies indicating the utility of our proposed framework as being used by business clients. We present studies based on the tweets crawled for specific commercial brands of three business customers – Sprint-Mobile, Crest Toothpaste, and Holiday-Inn hotel chain.

A. The Dataset

Our proposed framework as described in this paper is being used by business clients for monitoring and analyzing user feedback through Twitter as it has shown better potential in terms of its ability to capture the on-going events. In order to

Brands	# negative tweets	# of non-negative tweets
Sprint-Mobile	1,928	6,462
Crest	1,593	15,209
Holiday-Inn	2,339	41,577

TABLE I: DatasetStatistics

evaluate the capability of reporting the client-specific events of our event detection framework, we crawl twitter to gather tweets for three specific brands – Sprint-Mobile, Crest, and Holiday-Inn via the API provided by Twitter. The tweets were collected based on a keyword matching by using a set of client supply keywords. Some statistics about the tweets thus collected are reported in Table I.

B. Setting Up and Configuring the Framework

Our framework is implemented in Java and deployed on a server with Intel Core i5 CPU (2.40 GHz) and 8 GB memory. A separate server is employed to continuously crawl the tweets for the above mentioned three brands and save the crawled content into a MySQL database hosted locally. Thus, this crawling server is responsible for collecting tweets, while the framework on the deployment server is in charge of performing event detection over the tweets on the crawling server. As has been described before, text pre-processing on crawled tweets is performed using the Gate toolkit [2] and proprietary sentiment classifiers are used to identify the polarity of tweets.

In order to set up the batch version of the event detection framework, users could indicate a specific time range and a specific brand. The batch version then performs event detection over the brand specific tweets within that time range and reports the results to the user. Considering the suggestions given by domain experts, the data of the first 30 days is used for training, i.e., computing the the normal negative frequencies of a terms/phrases. The data for each of the subsequent days is used for testing, i.e., detecting terms/phrases with an abnormal negative frequency (too high comparing with the training data in the first 30 days). It is worth noticing that even the frequency of 1 for a term would be considered abnormal if the historical frequency for that term is 0. Thus, a frequency threshold (in this case, it is set as 5) is specified while determining if a given term is an anomalous event or not. Further, in order to restrict the number of detected events to be analyzed by the customer care agents, only the top 50 most frequent nouns/phrases are treated as the candidates of events which should draw agents' attention. All the parameters as discussed can be easily configured by the user depending upon his requirements through a graphical interface

The set up of the streaming version is more complex. Advised by the domain experts, the tweets of the first 30 days are used for training; only the top 3,000 nouns/phrases in the historical window (30 days) are maintained for detecting events. For every subsequent hour, the top 3,000 nouns/phrases would be updated based on the newly encountered tweets within that hour. In the event detection procedure, the tasks of "new" and "anomalous" event detection are executed every 24 hours: for the "new" event detection, if a specific event does not show in the last 30 days, and it appears more than 3 times within the latest 24 hours, we mark it as "new"; for the "anomalous" event detection, if a specific event appears more than 5 times within the latest 24 hours and is considered to be abnormal via Grubbs' test, it is marked as "anomalous". Note that the numbers 3 and 5 here are thresholds for "new" and "anomalous" event detections respectively. Since "anomalous" event detection has a more strict threshold than "new" event detection, an event could be "new" first and then "anomalous" later on. Again, these parameters can easily be adjusted based on client's requirements.

C. Detection Examples

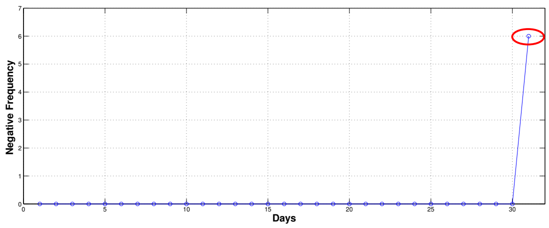
In this section, we first show two examples illustrating differences between "new" and "anomalous" events. Then a few "new" and "anomalous" events for the three brands are presented in Figure 5. Note that the core of detection concept of both batch and streaming versions are similar, so the detection results focus on the streaming version. The results for anomalous event detection as obtained by the batch version for Sprint-Mobile are shown in the Figure 6 as an example. In all the examples, event detection has been performed on tweets with negative sentiment.

In order to show the detection results of our proposed event detection framework, two example events are first presented in Figure 3 (for Holiday-Inn dataset) and Figure 4 (for Crest dataset) respectively. These two example events represent new (Figure 3) and anomalous(Figure 4) events, respectively. Both Figures are composed of two parts: the left part represents the changes of negative term frequency of the event within the 31 days; while the right part shows the associated tweets for the 31st day, reporting what people mentioned about this event on the 31st day. The combination of these two parts illustrates the root cause of raising this event as "new" or "anomalous".

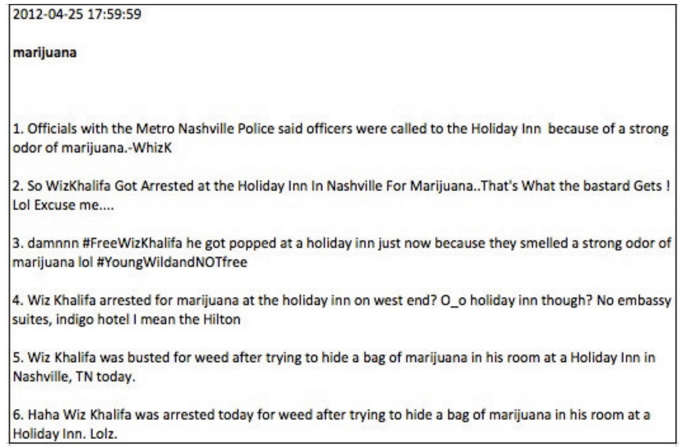
One can observe from the Figure 3 that the event "marijuana" was suddenly observed in many tweets about Holiday Inn because of the arrest of a celebrity in a Holiday Inn hotel during the 31st day of the dataset, and since it was never mentioned before, it was identified as a "new" event. On the other hand, we observe from Figure 4, that the event "advert" was observed continuously in tweets related to Crest. However, poorly received Crest advertisement caused a sudden increase in the number of Crest related tweets containing the term "advert" and this behavior being different from the historical pattern as observed for this event, resulted in it being labeled as an anomalous event.

In order to have a big picture about the detection results, those detected events via the streaming event detection for each of the three brands are reported in Figure 5. In this Figure, one can see that the detected events from 2012-04-21 to 2012-04-26 are listed, and those events illustrate the trending topics in Twitter about the particular brands. Through reading about these events and their associated tweets, users can quickly capture where the complaints come from.

In addition to the streaming event detection results, a set of detected events by the batched event detection for Sprint-Mobile are also presented in Figure 6 to show the capability of the batched processing.



((a)) Negative term frequency of each day



2012-04-25 17:59:59

marijuana

1. Officials with the Metro Nashville Police said officers were called to the Holiday Inn because of a strong odor of marijuana.-WhizK

2. So WizKhalifa Got Arrested at the Holiday Inn In Nashville For Marijuana..That's What the bastard Gets ! Lol Excuse me....

3. damnnn #FreeWizKhalifa he got popped at a holiday inn just now because they smelled a strong odor of marijuana lol #YoungWildandNOTfree

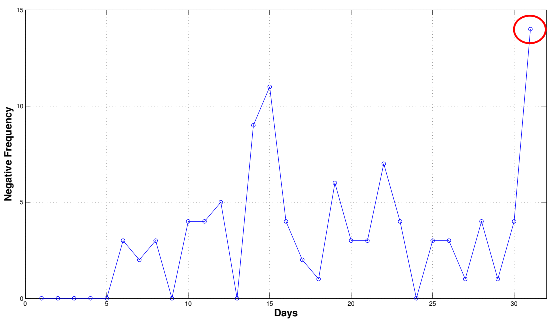
4. Wiz Khalifa arrested for marijuana at the holiday inn on west end? O_o holiday inn though? No embassy suites, indigo hotel I mean the Hilton

5. Wiz Khalifa was busted for weed after trying to hide a bag of marijuana in his room at a Holiday Inn in Nashville, TN today.

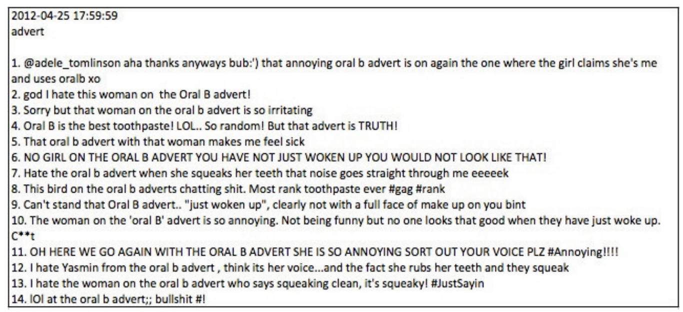
6. Haha Wiz Khalifa was arrested today for weed after trying to hide a bag of marijuana in his room at a Holiday Inn. Lolz.

((b)) The associated tweets of the 31st day

Fig. 3: A sample new event – Marijuana.



((a)) Negative term frequency of each day



2012-04-25 17:59:59
advert

1. @adele_tomlinson aha thanks anyways bub:') that annoying oral b advert is on again the one where the girl claims she's me and uses oralb xo
2. god I hate this woman on the Oral B advert!
3. Sorry but that woman on the oral b advert is so irritating
4. Oral B is the best toothpaste! LOL.. So random! But that advert is TRUTH!
5. That oral b advert with that woman makes me feel sick
6. NO GIRL ON THE ORAL B ADVERT YOU HAVE NOT JUST WOKEN UP YOU WOULD NOT LOOK LIKE THAT!
7. Hate the oral b advert when she squeaks her teeth that noise goes straight through me eeeeek
8. This bird on the oral b adverts chatting shit. Most rank toothpaste ever #gag #rank
9. Can't stand that Oral B advert.. "just woken up", clearly not with a full face of make up on you bint
10. The woman on the 'oral B' advert is so annoying. Not being funny but no one looks that good when they have just woke up. C**t
11. OH HERE WE GO AGAIN WITH THE ORAL B ADVERT SHE IS SO ANNOYING SORT OUT YOUR VOICE PLZ #Annoying!!!!
12. I hate Yasmin from the oral b advert , think its her voice...and the fact she rubs her teeth and they squeak
13. I hate the woman on the oral b advert who says squeaking clean, it's squeaky! #JustSayin
14. lOl at the oral b advert;; bullshit #!

((b)) The associated tweets of the 31st day

Fig. 4: A sample anomalous event – Advert.

V. Related work

The tremendous growth and popularity of social media platforms like Twitter, Facebook, etc. provides business organizations an opportunity to monitor the feedback from its customers, identify their problems and take corrective measures. Studies have shown that monitoring user feedback and communicating with them over social media can help boost a company's sales and reputation amongst consumers [5]. Lauschke and Ntoutsi [6] describe techniques for analyzing individual users and monitor changes in their interests over time. Our work addresses the problem of monitoring consumer feedback through social media platforms and use this to identify possible issues face by the consumers. We have framed the problem as an event detection problem where the event corresponds to an issue faced by consumers. Event detection is a popular problem and has been applied to many real world applications, such as surveillance[7], scientific discovery, and fault detection. It has been used to detect disease outbreaks before the situation turns to an endemic [11], thus saving a large number of human lives and money. As another example, by applying the event detection techniques for environment monitoring [3], anomalies in environment sensor data have been discovered and corrected. Further, event detection techniques have also been used in disaster management scenarios such as earthquake detection by monitoring tweets [10], illustrating the speed of information diffusion and propagation through popular social media platforms, and hence, the motivation to monitor customer feedback through social media platforms.

VI. Conclusions and future work

In this paper, we describe our attempts at designing a platform to help business organizations automatically monitor consumer feedback through social media documents and identify issues faced by consumers, thus enabling organizations to take corrective measures in a timely fashion. We illustrate the utility of the proposed framework by describing a few case studies describing how our proposed framework is being used by three different clients. Based on our interaction with the clients using

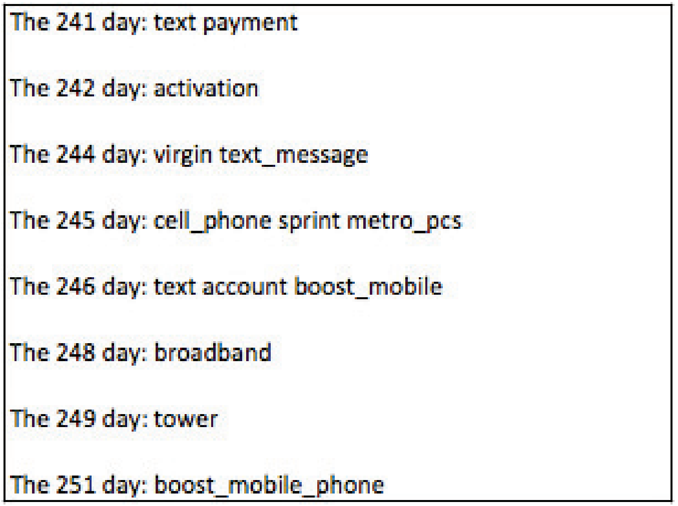


Fig. 6: Batched event detection results for Sprint-Mobile.

the framework, there are a lot of areas offering scope for further research and development. First, for a particular event of a brand, the negative term frequency may be contributed by users from different geo locations. Identifying the geo locations of those users can help the clients figure out the root cause of the issues and take area specific corrective measures. Second, not every event may be equally important for the client and hence, a mechanism for measuring the importance of events based on client requirements is necessary. Third, preprocessing social media documents is a hard problem as they usually contain a lot of noise (abbreviations, spelling variations and mistakes, etc.). Thus, new NLP tools and algorithms tailored to social media text need to be developed.

REFERENCES

[1] N. G. Barnes, "Exploring the link between customer care and brand reputation in the age of social media," *S. f. NC Research (Ed.): Society for New Communication Research*, 2008.

[2] H. Cunningham, "Gate, a general architecture for text engineering," *Computers and the Humanities*, vol. 36, no. 2, pp. 223–254, 2002.

[3] E. Dereszynski, "A probabilistic model for anomaly detection in remote sensor streams," 2007.

[4] N. Kanhabua, S. Romano, and A. Stewart, "Identifying relevant temporal expressions for real-world events," in *Proceedings of The SIGIR 2012 Workshop on Time-aware Information Access, Portland, Oregon*, 2012.

[5] J. H. Kietzmann, K. Hermkens, I. P. McCarthy, and B. S. Silvestre, "Social media? get serious! understanding the functional building blocks of social media," *Business Horizons*, vol. 54, no. 3, pp. 241–251, 2011.

[6] C. Lauschke and E. Ntoutsi, "Monitoring user evolution in twitter," in *Workshop on Business Applications of Social Network Analysis, collocated with ASONAM, 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE, 2012, pp. 972–977.

[7] M. H. Y. Liao, D.-Y. C. D.-Y. Chen, C.-W. S. C.-W. Sua, and H.-R. T. H.-R. Tyan, "Real-time event detection and its application to surveillance systems," pp. 509–512, 2006. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1692634

[8] W. G. Mangold and D. J. Faulds, "Social media: The new hybrid element of the promotion mix," *Business horizons*, vol. 52, no. 4, pp. 357–365, 2009.

[9] N. I. of Standards, T. (US), C. Croarkin, P. Tobias, and C. Zey, *Engineering statistics handbook*. The Institute, 2001.

[10] T. Sakaki, M. Okazaki, and Y. Matsuo, *Earthquake shakes Twitter users: real-time event detection by social sensors*. ACM, 2010, pp. 851–860. [Online]. Available: http://dl.acm.org/citation.cfm?id=1772690.1772777

[11] M. Wagner, F. Tsui, J. Espino, V. Dato, D. Sittig, R. Caruana, L. McGinnis, D. Deerfield, M. Druzdzel, and D. Fridsma, "The emerging science of very early detection of disease outbreaks," *Journal of Public Health Management and Practice*, vol. 7, no. 6, pp. 51–59, 2001.

	Sprint-Mobile	Crest	Holiday-Inn
New Events	2012-04-21 virgin_mobile_peep belmont virgin_mobile_network 2012-04-22 cripple upgrade phone_charge log data_access android_phone 2012-04-23 boon roaming 2012-04-24 screen instruction facebook_account felon technology trip phone_ring tyler 2012-04-25 tyler peak hood insanity virgin_mobile_web foothill lake_lanier contract_t-mobile paris phone_instruction 2012-04-26 dispute cellular calling_card moblie cricket-phone textgram	2012-04-21 match food mouth_kyle 2012-04-22 package yolo crest_fight_plaque tesco_finest crack kit wax crest_mouth_wash crest_sry 2012-04-23 reservation competition commercial mystery directioner mango hoop hammersmith_crest tone bass earth sudden scum drink shop beauty 2012-04-24 enamel monger rembrandt_teeth_whitening_kit blend apple tea cracked_crest powerbrush buck rinse barett clip shot uble roomie lip flavor_toothpaste marc tongue 2012-04-25 strip_sample crest_whitening ulta_purchase lecture germ butter amber shield pro_expert sperm crest_kids_sparkle confession intention doctor_dental stank deodorant scope_whiting_mouth_wash issue toothpaste_ad ache braun_electric_toothbrush 2012-04-26 travel_size laurel bright scope_mouth_wash copenhagen design bottom mojo crest_whitening_advanced_toothpaste	2012-04-21 sale toy realtalk holiday_inn_johnstown themosthighgod bed_bug_bite evictionpartyinmyroom bean holiday_inn_paris_bastille sun yolo bevvy bill gurnee wycombe holiday_inn_madness hock landscape voyage peacock crawfish_festival louse holiday_inn_club_vacation_orange_resort 2012-04-22 pun trk_chilling chelmsford brand_holiday_inn marathon anw logo security train_station maple travel_lodge bath hotel_bathtub holiday_inn_tub 2012-04-23 brentwood marketing_mail catering holiday_inn_broadband joburg bedford staff_response broadband obligation sheet recipe steal metal pilot pavement stadium blanket leicester_holiday_inn prob_holiday_inn dance canuck pool_deck polynesian estate_license 2012-04-24 pet robe slipper umbrella budget holiday_inn_bandwidth shite_internet nozzle 2012-04-25 holiday_inn_priority_club wiz scheme kidnapping cameron weed wiz_khalifa celebrity_news_excitement celebrity shout marijuana_news marijuananews jail sister rapper nerd 2012-04-26 mouse snoop metro_police hairdryer nashville nashville_holiday_inn priority_club_card clark
Anomalous Events	2012-04-21 android_smartphone price pcd_venture 2012-04-22 iphone 2012-04-23 iphone android_smartphone price pcd_venture virgin_mobile 2012-04-24 android_smartphone price pcd_venture virgin_mobile text 2012-04-25 2012-04-26	2012-04-21 fashionista spring_color_trend 2012-04-22 2012-04-23 competition 2012-04-24 competition enamel monger rembrandt_teeth_whitening_kit powerbrush teeth 2012-04-25 teeth powerbrush competition crest_pro_health advert 2012-04-26 advert crest_pro_health	2012-04-21 breakfast motel 2012-04-22 motel hotel 2012-04-23 2012-04-24 holiday_inn_express molina doctor 2012-04-25 motel holiday_inn_express molina hotel doctor wiz_khalifa weed wiz 2012-04-26 wiz_khalifa weed wiz chillin

Fig. 5: Detected events for three clients.