

RESEARCH

Open Access



# Localized twitter opinion mining using sentiment analysis

Syed Akib Anwar Hridoy, M. Tahmid Ekram, Mohammad Samiul Islam, Faysal Ahmed and Rashedur M. Rahman\*

\*Correspondence:  
rashedur@northsouth.edu  
Electrical and Computer  
Engineering Department,  
North South University,  
Bashundhara, Dhaka,  
Bangladesh

## Abstract

Analysis of public information from social media could yield interesting results and insights into the world of public opinions about almost any product, service or personality. Social network data is one of the most effective and accurate indicators of public sentiment. In this paper we have discussed a methodology which allows utilization and interpretation of twitter data to determine public opinions. Analysis was done on tweets about the iPhone 6. Feature specific popularities and male–female specific analysis has been included. Mixed opinions were found but general consistency with outside reviews and comments was observed.

**Keywords:** Data mining, Natural language processing, SNLP, SentiWord, Rapid Miner

## Background

Sentiment analysis technique is an effective means of discovering public opinions. Various companies often use online or paper based surveys to collect customer comments. Due to the emergence of social networking sites and applications, people tend to comment on their facebook or tweet profile. Therefore the paper based approach is not an efficient approach. Only a very small customer base can be reached and there is no guarantee that their answers in the survey are honest or not. Here social media comes into play. Facebook, Twitter and all other social media sites are full of people's opinions about products/services they use, comments about popular personalities and much more. Hence mining opinions about various subject matters from social media is a much more innovative approach for market analysis. A lot of research has been done on opinion mining from social media, most of which focuses on people's sentiment towards various topics. But analyzing social media data in this manner gives a much generalized idea. To make it more specific, sentiment analysis can be performed on social media data from explicit locations. Our approach is to find the sentiments in specific locations. This will allow companies to focus their marketing expenditures on areas where sentiment is low, while maintaining minimum advertisement in areas of high popularity.

In this research we have analyzed a large data set from which we tried to determine the popularity of a given product in several locations. In order to do this we analyzed tweets from Twitter. Tweets are a reliable source of information mainly because people tweet about anything and everything they do including buying new products and reviewing

them. Besides, all tweets contain hash tags which make identifying relevant tweets a simple task. A number of research works has already been done on twitter data. Most of which mainly demonstrates how useful this information is to predict various outcomes. Our current research deals with outcome prediction and explores localized outcomes.

We collected data using the Twitter public API which allows developers to extract tweets from twitter programmatically.

The collected data, because of the random and casual nature of tweeting, need to be filtered to remove unnecessary information. Filtering out these and other problematic tweets such as redundant ones, and ones with no proper sentences was done next.

As the preprocessing phase was done in certain extent it was possible to guarantee that analyzing these filtered tweets will give reliable results. Twitter does not provide the gender as a query parameter so it is not possible to obtain the gender of a user from his or her tweets. It turned out that twitter does not ask for user gender while opening an account so that information is seemingly unavailable. We used an AI tool which can be used to accurately deduce gender from a person's user name called NamSor.

After completion of the analysis phase, experimental results were presented. A variety of results was attainable from the available data, so we decided to present only those which accurately reflected the sentiment of the people towards the product. Several nationwide metrics, city by city metrics and gender separated metrics for individual cities had been discussed.

The rest of the paper is organized as follows. Previous research that has been done using twitter data mainly to determine sentiment has been discussed in "[Related work](#)". The entire data extraction procedure has been explained in "[Data extraction](#)", followed by steps taken to filter and preprocess the tweets in "[Data preprocessing](#)". The implementation of the sentiment analysis program and related applications are discussed in "[Implementation](#)". The results of the experiment have been visualized graphically and also interpreted. The inferences made along with supporting literature has been presented in "[Result](#)". How the methodology could be applied to non-technical areas in order to determine other sorts of opinions or sentiments has been discussed in "[Other applications](#)". Finally "[Conclusion](#)" draws the conclusion and "[Limitations](#)" and "[Future work](#)" discuss limitations and future work respectively.

## **Related work**

Nithish et al. (2013) mainly focused on market reaction using sentiment analysis. The research focused on smart phone domain. They intended to find out the explanation of what influences a device's rating. As twitter is one of the most widely used social media for microblogging, they used tweets related to the product to obtain dataset. People have their own opinion regarding a specific product either good or bad or mixed review. So they clustered the tweets using NLP to determine positive or negative or neutral feedback. They created an ontology to get the meaning of tweet out of the sentence. Ontology is a domain model which includes all the prevalent features as well as the relationships between them. Initial ontology was created using data from online store. Tweets were retrieved using keywords. Sentiment analysis was then performed. The ontology was updated using scores from sentiment analysis. Various queries were performed on the

updated ontology. From this research we learn that twitter data is a good indicator of public opinion and it can be used to accurately determine sentiment.

For this data mining research Khanaferov et al. (2014) first collected unstructured data from one of the most popular social networks, Twitter. Their goal was to demonstrate a practical approach to solve an alarming healthcare issue through a computational approach centered on mining useful patterns out of public data. The main purpose was to demonstrate the power of mining unstructured data from an unlikely source of data. They selected healthcare informatics to demonstrate the significance of data for a complex domain. NLP is an active research field and many tools have been developed by prominent organizations that allow for semantic processing. Main focus was on a subset of NLP for the purposes of this study. First step was to clean up the data by removing records with missing attributes. In a subsequent step, user location strings were mapped into usable latitude and longitude data. Each keyword id represents a keyword object stored in the keywords table. Once the data was filtered a standardization phase was followed. In this phase the cleaned data was normalized using mathematical functions. A density based clustering algorithm was selected. DBSCAN is a type of density based clustering algorithm. The output of the DBSCAN algorithm is a set of clusters. To visualize the clusters created by the DBSCAN algorithm, the results were mapped onto Google Maps using Google Map API 3.0. The localized results were consistent. Continents east of the Atlantic showed positive sentiment while that west of the Atlantic showed negative sentiment towards the obese, which is true. Here we can deduce that by using keywords we can obtain topic specific tweets. This makes analyzing the data easier because we can be sure that all tweets are from the same domain.

The main aim of this research by Kim et al. (2013) was to detect short period trends on twitter. Generally this refers to events or holidays or anything similar which lasts for a while and then loses activity. A problem that was encountered was that simply counting the word frequency was not enough to discover a trending topic. This is because commonly used words such as 'love', 'like' are very common in all tweets and will obviously have a high frequency no matter what set of tweets is analyzed. The approach used by the authors involved plotting the tweet's frequency as a function of time. This resulted in a very helpful pattern where commonly used words had a very much constant frequency throughout the time period being considered but certain keywords showed spikes during certain times. The two events that were analyzed were Easter and weather patterns. The results were very clear as keywords related to Easter spiked on the day of Easter and slowly dropped down in the next couple of days. Similarly areas from weather related tweets were obtained showed sentimental consistency with the actual weather situation in that given area. The results were extremely consistent with the real world events as the output of the weather patterns was accurate as they matched with all the weather forecasts. So the way people are commenting on twitter about the weather was a good indicator of the actual weather turn out again justifying the accuracy of information prediction using twitter. This study shows people talk about certain events on twitter and it turns out that these discussions reflect actual events in those places. So if people in a given location are discussing a product on twitter, their opinions on twitter reflect the actual sentiment in that area.

Akhtar (2014) used various social network analysis tools, for example, Gephi, Networkx, IGraph and Pajek and reported comparative results on efficiency, visualization and graph features. At the end the authors concluded that IGraph outperformed other tools in processing complex and large network.

Ostrowski (2012) discussed a method by which social network data can be analyzed to find trends and people of power or influence within a given community or network. The idea was to represent an entire social network subset from twitter in the form of a graph where each node represents a person and the edges represent some form of connection that exists among people. So generally nodes with more edges will usually be identified as influential. They used the 'reply' concept in twitter meaning they analyzed how many people comment on or reply to a tweet as an indicator of how influential the person is. This was also done over a long period of time to give more consistency to the results. After analysis of a social network graph generated from twitter data based on three major mobile operating systems, the research concluded that the graph metrics were consistent with the real world information showing that social network data is good for trend and influence identification.

Cho et al. (2014) focused on finding out the brand image of a particular brand. The work was done on Korean tweets. Rather than using general or publicly available sentiment analysis schemes, these tweets were first analyzed using a morpheme analyzer (a morpheme is the smallest meaningful unit of a language) in order to find which morphemes are important to the sentiment analysis and use them to construct sentiment dictionaries which will only focus on the different brands alone. This allows the sentiment analysis phase to work with the brand names grammatically, something which is not possible with commercial default tools. The temporal and spatial changes to a brand were determined. Samsung product was analyzed to see how the sentiment was on different states in the country, this provided spatial change. While two major political parties were analyzed over a period of time to get temporal change of sentiment. Both product and non-product subject matters have been studied in this research. This shows that sentiment analysis is versatile and applicable in any field.

Servia-Rodriguez et al. (2013) mainly focused on extracting user interest by analyzing the user's tweets. The tweets were first analyzed using natural language processing tools in order to characterize the person's tweets. Main focus was on nouns because these refer to interests directly. All available tweets from one user were analyzed and then clustered based on how close they were logically or semantically. The clustering was done using three tag clustering algorithms called PAM, Affinity Propagation and UPGMA. A comparison was made on the quality of the clusters produced by these algorithms with unsupervised inputs. The experiment resulted in the UPGMA being the best option for such clustering analysis tasks. Therefore, it is possible to filter tweets beforehand by using natural language processing so that tweets which do not express opinions are removed before experimental analysis.

Ostrowski (2013) discusses how application of semantic filtering on twitter data can be used to determine topics trending currently. After a stage of empirical filtering where the tweets go through very rudimentary filtering, a knowledge base is developed by analyzing the tweets using machine learning algorithms. Trend plots were then taken for periods of time which showed periods of spikes and drops. This data was consistent with

the data from Google Trends, which is Google's proprietary trend identification tool. By comparing the experimental results with that from Google Trends the authors were able to establish the accuracy of their method.

### Data extraction

Twitter tweets were used as a data source. It is possible to extract tweets in a large scale from Twitter using the twitter public API that they provide. In our case we used the "twitteroauth" version of the public API by Williams (2012). This version has been implemented in PHP and can be run directly on the local host or on web servers. The query could contain several parameters. Twitter provides a large set of filtering parameters so that a well-defined set of tweets can be obtained. Once the query has been constructed it can be ran by the API and all relevant twitter data will be provided as output in the browser. This data was directly inserted into a MySQL database for the use later on.

Each record or tweet that is obtained contains several types of information like user name, tweet id, text etc. But out of those only the text and tweet id were useful to us. Initially the twitter API allowed tweet locations in the form of latitude and longitude to be available with every tweet were the user has made his/her location public. But due to security issues and user complaints this was stopped in 2012. This means that the geographical location from where the tweet was created is not available with the tweet. What twitter does allow on the other hand is the use of location as a filtering parameter in the main query. So in compliance with this restriction we had to extract tweets based on a fixed set of locations.

For our research we decided to focus on one nation, USA. We extracted tweets from seven major cities in the USA. The choice of location is very limited mainly due to data availability and language constraints. We decided to go with data from New York, Los Angeles, Boston, Chicago, Dallas, San Francisco and Philadelphia for the experiments.

Each major city has a city center, the latitude and longitude that was used to define the city itself. The radius of coverage was chosen based on approximate measure obtained from using free map tools by Viklund (2015). The radius was picked in such a way so that major parts of the city were covered. Even if a bit of excess was covered it does not really matter as those areas are generally very lightly populated and will not give results anyways. The latitude, longitude and radius are all values assigned to the 'locations' parameter in the query build. So now we have multiple data sets each obtained from a different city.

The product that we chose to analyze was the iPhone 6. Even though it is possible to analyze any product's popularity using the defined method, the availability of data was an important issue. At the time of this research the only electronic product trending on twitter was the iPhone 6. Meaning that a reasonable amount of data about this device was available. But our method can be used to obtain results for any product given that a good amount of data about it is available on Twitter. So only the tweets which contained the term 'iPhone 6' in them were obtained. As we also decided to determine which feature of the iPhone 6 was most or least popular the query was enhanced using a few keywords to obtain feature specific tweets. An example would be 'iPhone 6 battery'. This query parameter will cause the API to return only tweets which contain both iPhone 6 and battery terms together which results in tweets about the battery performance of

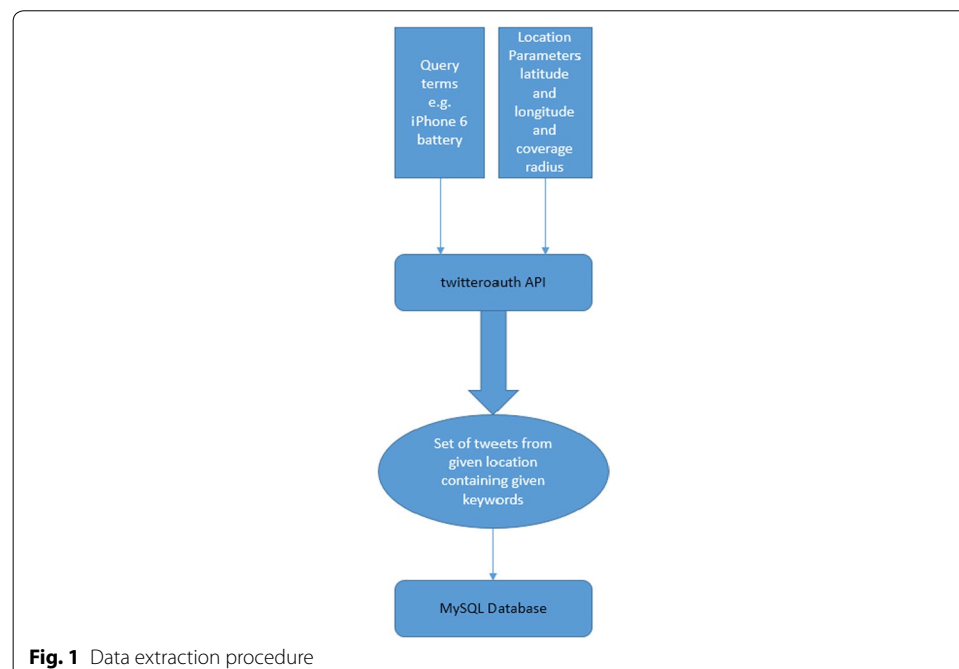
the iPhone 6. Other keywords used were “camera”, “iOS”, “iTunes”, “screen”, “sound”, and “touch”. For each tweet, the user name, tweet text, location were extracted.

Originally through the twitter API we collected 940 tweets. Out of those tweets 530 were from male users and 410 were from female users. Number of tweets from New York, Los Angeles, Boston, Chicago, Dallas, San Francisco and Philadelphia are 182, 89, 103, 143, 156, 138 and 129 respectively. But once NLP filtering was applied 442 tweets were left. The rest of the tweets were not useful in any way for sentiment analysis. A flow chart illustrating the data extraction process is shown in Fig. 1.

### Data preprocessing

The data obtained from the API obviously contains a lot of non-relevant data. Very basic and rudimentary cleanup was performed using Java. Arbitrary characters and other useless information in a tweet were filtered out before further analysis.

In order to filter out these useless data we mainly used the Stanford Natural Language Processing tool by The Stanford NLP Group (SNLP Group 2015) which is an open source natural language processing tool developed by Stanford University. This tool was used because it gives the grammatical relations between the words in a sentence as output. According to advanced linguistics several such relations are available in the English language. But not all relations are useful in general natural language research. So SNLP has 50 predefined relations which they call dependencies. These dependencies are listed and explained in the Stanford Type Dependencies Manual (SNLP Manual 2015). The reason 50 dependencies are defined in the SNLP is because these are the only word relations which are useful to information analysts, even though linguistics defines several other word relations within a sentence. Out of these 50 dependencies we chose three which will be useful to us. These are **nsubj**, **amod**, **dobj**. These are relations that are



required to identify the tweets that will be useful and contain meaningful information. Further filtering with even more relations does not aid the results in anyway.

The **nsubj** relation is used to find relations between nouns and adjectives or verbs which are complementing the noun in a sentence. This is extremely important because it gives an idea whether or not the sentence is in anyway complementing a noun or not. An example of the **nsubj** relation would be as follows.

*"My iPhone 6 camera is awesome!"*

For the above sentence we will obtain several relations including **nsubj (camera, awesome)**. This relation shows that the 'camera' noun has been linked with the 'awesome' adjective meaning that this tweet will be useful for the sentiment analysis.

The **amod** relation is the adjectival modifier. This is used to find any adjectives that are used in a sentence to modify a noun phrase. An example would be.

*"Got the new gold iPhone 6, feeling great!!"*

For the above tweet the adjectival modifier would be **amod (iPhone 6, gold)**, meaning that gold is modifying the noun phrase iPhone 6.

The **doobj** relation is the direct object, which is used to identify direct objects that a verb is referring to in a sentence. An example would be

*"Love the camera of iPhone 6!"*

For the above tweet the direct object relation would be **doobj (love, camera)** and also **doobj (love, iPhone 6)**. This is again crucial to the filtering process because these verbs will affect the sentiment analysis significantly.

Now such relations can be obtained for every tweet. The way we filtered out the unnecessary tweets is by removing those which do not result in a desired dependency and do not have at least one of the key words which have been fixed beforehand, within the desired dependency. So for a tweet to be valid and pass the preprocessing phase it must contain at least one dependency from the three dependencies above and this dependency must contain at least one keyword from the list of prefixed keywords. The fixed keywords basically contain words which must appear in the tweet in order to make it useful. For example iPhone, battery, screen, camera, sound etc. are a few such words.

Let  $t_1$  be a tweet from the set of tweets  $T$ .

If  $t_1$  contains

**nsubj ( $n_1, n_2$ ) V amod ( $n_3, n_4$ ) V doobj ( $n_5, n_6$ )**

Where at least one parameter,  $n_i$ , of the valid relations contains a keyword from the predefined list, then that tweet is said to be valid and is moved to the set of filtered tweets.

This is necessary because only extracting tweets with the word 'iPhone' will result in a large amount of unnecessary tweets an example of which would be,

*"Taking pics on my iPhone 6"*

The above tweet must be filtered out because it does not in any way refer to the quality or performance of the device. Rather it only showcases how the user is utilizing the



device for his/her personal use. This tweet will not contribute in any way to the sentiment score calculations and hence is not used in next phase.

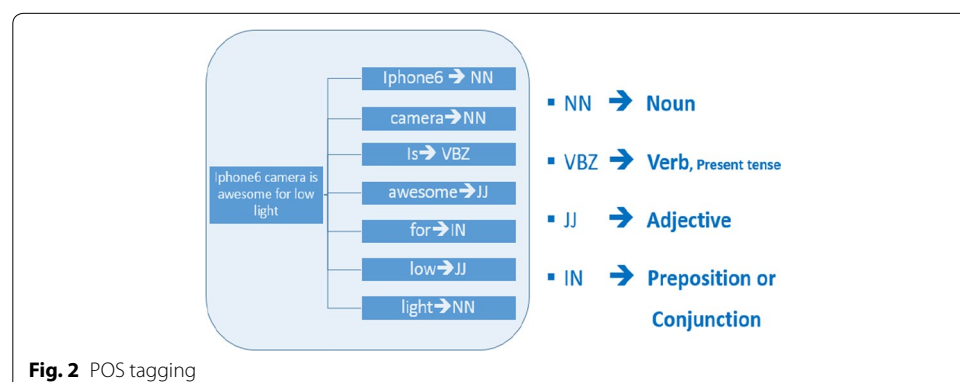
Originally through the twitter API we collected 940 tweets. But once NLP filtering was applied 442 tweets were left. The rest of the tweets were not useful in any way for sentiment analysis. Generally simple queries would normally return a lot of tweets. With a query which only contains the term iPhone 6 it is possible to get 1500 tweets or more in one run. But as more parameters are added to the query number of tweets returned is reduced. As we used location parameter and two search parameters, one for iPhone and the other for a feature, the number of tweets that we obtained was lower.

## Implementation

In order to assess the sentiment which is present in the tweet a numeric metric is required. This has been done using the tool SentiWordNet (2015), which comes bundled with the SNLP. What SentiWord does is that it takes a word and also the part of speech that a word has in a given sentence. Using the combination of part of speech and the word itself SentiWord gives it a numeric score between  $-1$  and  $1$  where lower value refers to more negative sentiment and higher value refers to higher sentiment. As a tweet text consists of a few words we can take the SentiWord score for each of those words and then sum them up to get a numeric score for each tweet.

Another issue here is that SentiWord does not recognize sentences; it only takes words and their corresponding part of speech as input. The part of speech the word will have will depend completely on the sentence itself. So a way has to be devised to map each word in the sentence to its corresponding part of speech. This was done using Parts of Speech tag extraction. This is also bundled with the SNLP and is used to identify the parts of speech a word has within a given sentence. So each tweet must first be analyzed using the POS tagger which will separate the tweet into individual words and assign a part of speech to it. This is required because by only assessing the word itself it is not possible to determine any sort of opinion, what part the given word plays within a sentence is always defined by the part of speech it using. Figure 2 illustrates the POS tagging process.

In order to map or normalize the POS tags assigned by the POS tagger we had to implement a custom program. Knowing that SentiWord only recognizes nouns, adjectives, adverbs and verbs, any parts of speech other than these three had to be mapped





to any one of these. An example of the mapping convention would be that if a word is assigned the VBZ tag, which stands for verb in present tense, it will be assigned the Verb tag by the mapper.

This set of words along with their normalized POS tags are then sent to SentiWord and the sentiment for each word is calculated and then the individual numeric sentiments are added to obtain a final score for the tweet. A complete example for a given sample tweet is shown below,

*“iphone6 camera is awesome for low light”*

The first column of Table 1 contains each of the words in the given tweet. The second column contains the corresponding parts of speech tag assigned by POS tagger. The third column contains the normalized POS tag which was mapped by the Mapping class. The fourth column is the SentiWord score. 0 value means that the word does not affect the sentence in any way. The total score is the sum of all the individual scores.

In this way we can obtain a sentiment score for each tweet. Now we will sum up the sentiment scores for all the tweets in a given location, such as a city, then divide the sum by the total number of tweets within that city which will give us an average score. This score will be the indicator of the people's sentiment towards the product in that given location. It is given in Eq. (1)

$$Score(location_j) = \frac{\sum_{i=1}^n SentiScore_i}{n} \quad (1)$$

where,  $n$  = total number of tweets,  $SentiScore_i$  = SentiWord score for each tweet,  $Location_j$  = Refers to one particular city.

As the scores obtained in this way do not follow any scale or are not within a given range it was necessary to normalize these scores in order to obtain fixed sentiment grades for the tweets. We adopted an approach similar to the normalizing process used by Nithish et al. (2013). Doing this allows any value within a given range to be assigned a sentiment which has been predefined for that range. The total score is the sum of all the individual scores and is normalized within  $-1$  to  $1$ . The normalization model is defined in Table 2.

In order to obtain the users' genders from the tweets we utilized a tool called NamSor (2015) which is a data mining tool offered as an independent product and also as an extension in Rapid Miner. We used the Rapid Miner extension for our gender

**Table 1 Scoring example**

Word	POS tag	Normalized POS	Score
iphone6	JJ	a	0.0
Camera	NN	n	0.0
Is	VBZ	v	0.0
Awesome	JJ	a	0.75
For	IN	Null	0.0
Low	JJ	a	-0.253290069096329
Light	NN	n	0.0568470433374004
Total score			0.5529180364277676

**Table 2 Normalization model**

Sentiment score range	Assigned sentiment
Score $\leq -0.5$	Worst
$-0.5 < \text{Score} \leq 0$	Bad
Score = 0	Neutral
$0 < \text{Score} \leq 0.5$	Good
Score $\geq 0.5$	Excellent

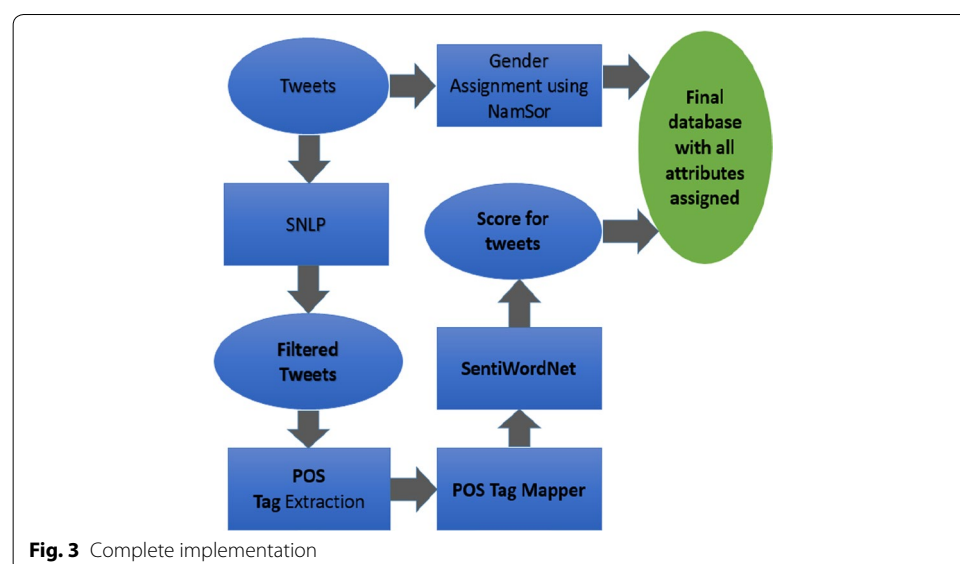
classification task. Once the filtered tweets were scored and placed into MySQL database, the database was exported into Rapid Miner and then the NamSor extension was applied to the database. The set of genders returned by NamSor was then inserted into the database for each corresponding tweet.

To make the result analysis phase bit simpler we integrated all the tweets into one single table with tweet text, user names, gender, feature, sentiment, and location as attributes. Originally the tables were separated with tweets from each city having their own tables. So running standard SQL queries on this table gives all the statistical results as required. The entire methodology has been illustrated in Fig. 3.

## Result

To properly understand the trends and variations in sentiments various comparisons were made. The comparisons started at a national level and then became more detailed by the introduction of cities and genders. A total of eight comparisons were made to illustrate the sentiment trends. These are as follows:

1. *National Average Sentiment* Sentiments inclusive of all cities and genders. It gives a general overview.
2. *National Feature Average Score* Average score inclusive of all cities but grouped by features. It gives general view of sentiment towards iPhone 6 features.



3. *National Male/Female Average Score* Average scores inclusive of all cities and features grouped by gender.
4. *National Male/Female Feature Average Score* Average scores inclusive of all cities grouped by gender and features individually.
5. *Average Score per City* Average sentiment score for the individual city.
6. *Male/Female Sentiment per city* Sentiment for each city grouped by gender.
7. *Feature Average Score per city* Average score per city grouped by feature.
8. *Male/Female Feature Average per city* Sentiment score for each city grouped by gender for each individual feature. This is a very important comparison because it involves all the variables, specific location, gender and feature.

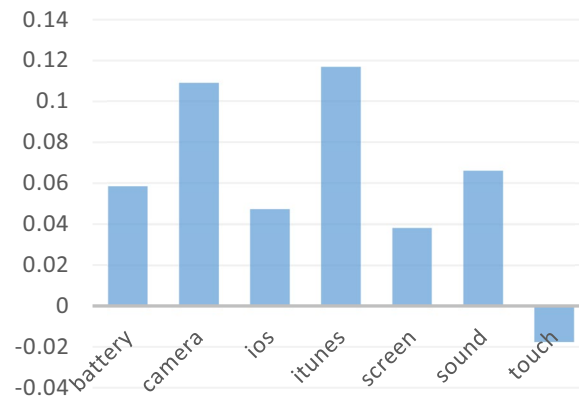
All of the comparisons have been illustrated using graphs for easy understanding and comparability. The average scores taken are standard averages, not weighted. As the scores were normalized beforehand it was unnecessary to renormalize the average scores. The sentiment percentages were found using the proportion of tweets having a given sentiment among all the other tweets.

As seen in Fig. 4 over 60 % people thought of the iPhone 6 as a good mobile device which is true because of the positive reviews the phone has received throughout. Many popular websites like the one by Beavis (2015) have reviewed the product as being of top quality. According to such reviews the positive sentiment towards the device is clear. As the results of the experiment are consistent with these reviews the research demonstrates that the methodology presented is effective in accurately determining sentiment. Generally excellent sentiments are difficult to obtain, mainly because this would require vast amount of tweets to contain words which have very highly positive SentiWord scores. This is generally never the case because when people express their thoughts about a device they tend to stick with simple descriptive terms. But even then the high level of good sentiment is still an accurate indicator.

When it came to the features of the iPhone 6, as illustrated in Fig. 5, camera and iTunes were two of the most popular features. The iPhone 6 camera has been highly praised and also been claimed to have an image quality similar to DSLR cameras. Also the iTunes music system is the most popular and user friendly music and content management system. This explains why these have high positive sentiments. The reason behind very low sentiment for the iPhone screen and negative sentiment for the touch is the 'bend' issue many users faced. After the first few complaints came in that the new iPhone would



**Fig. 4** National sentiment as percentage



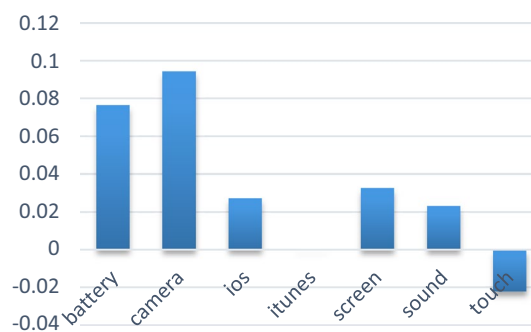
**Fig. 5** National feature average

bend inside the pocket like flexible metal or plastic these raised a storm of discussions on social media, regarding whether or not this was an unacceptable manufacturing failure from Apple's point of view, or some new sort of feature as discussed in the article by Rubin (2014).

Even though Apple claimed that this was a new way of making the phone more durable and impact resistant, the users were not satisfied with the claims and went on to protest the bending issue. Many cases reported that the phone bent so much that it didn't go back into shape or the touch stopped working or the screen glass cover simply broke. This issue alone created largely negative sentiments about the screen and touch features of the iPhone 6.

For the nationwide scores based on genders, mixed results were obtained. First noticeable pattern is with the screen and touch sentiments. This is consistent with the national feature sentiments for both male and female users. Therefore making more impact on the device's bending issue which the users did not take very positively.

Looking at Fig. 6 it is seen that male users generally have greater positive sentiment towards the software features of the iPhone 6. A very high sentiment for iOS and iTunes can be seen. A general conclusion that can be drawn from this is that male users tend to use a good number of apps and play games on their devices. This is again supported by



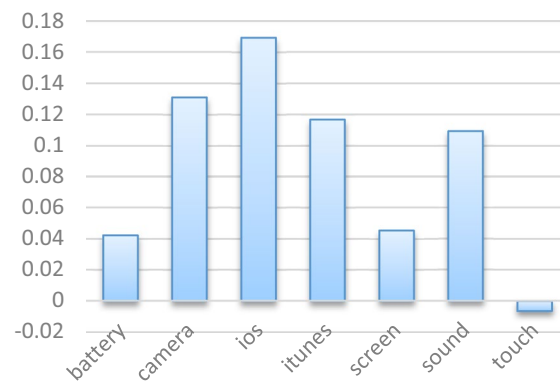
**Fig. 6** National male feature average

the low positive sentiments for the battery as heavy gaming and application use drains the battery easily.

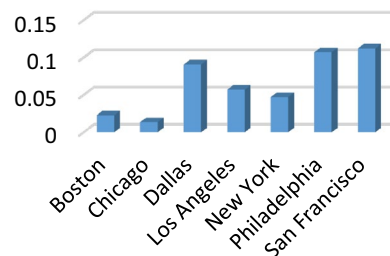
In contrast Fig. 7 shows a very different scenario for the female users of the new iPhone. The camera has received the highest sentiment amongst female users. Females generally have a higher tendency to take pictures and post them on twitter, this is generally a well-known fact. This coupled with the iPhone's already well reputed camera, supports the high sentiment. Also another point that can be deduced is that female users do not use as many apps or play games like male users. This can be deduced based on very low positive reference towards the iOS system and zero reference of the iTunes in any of the tweets collected from female users. It is further backed up by the high positive sentiment for the battery which is again contrasting the male scenario.

Looking at Fig. 8, the cities of San Francisco and Philadelphia have most positive sentiment scores with Dallas following on close by. Los Angeles and New York have medium positive sentiments while the cities of Boston and Chicago have lower positive sentiments. High popularity of the iPhone 6 in San Francisco can be explained by the presence of the Silicon Valley in the San Francisco bay area. This area houses some of the biggest technology companies in the world. So more tech savvy people reside in those areas and are clearly quite fond of the iPhone 6.

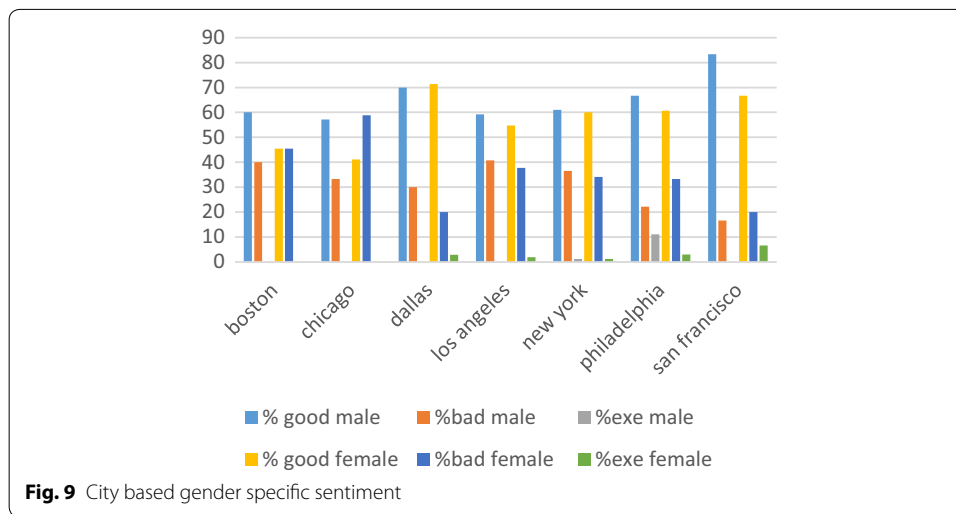
San Francisco shows highest positive sentiment scores for both female and male users, followed closely by Dallas and then by Philadelphia and New York. Philadelphia and New York show, according to Fig. 9, the broadest spectrum of tweets, with scores for all



**Fig. 7** National female feature average

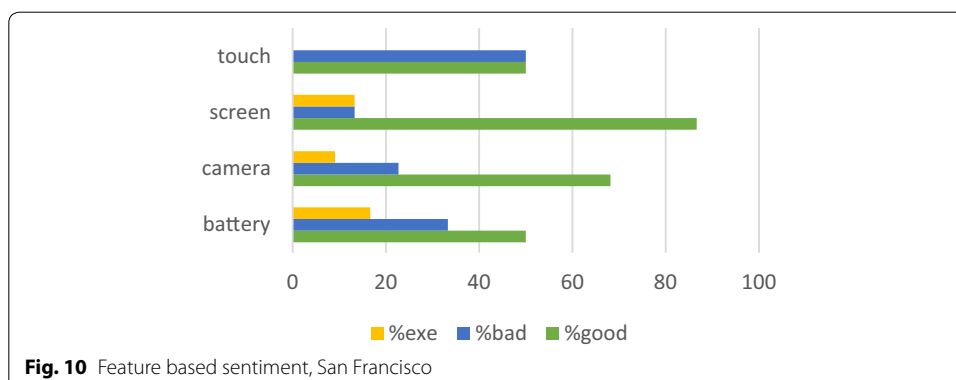


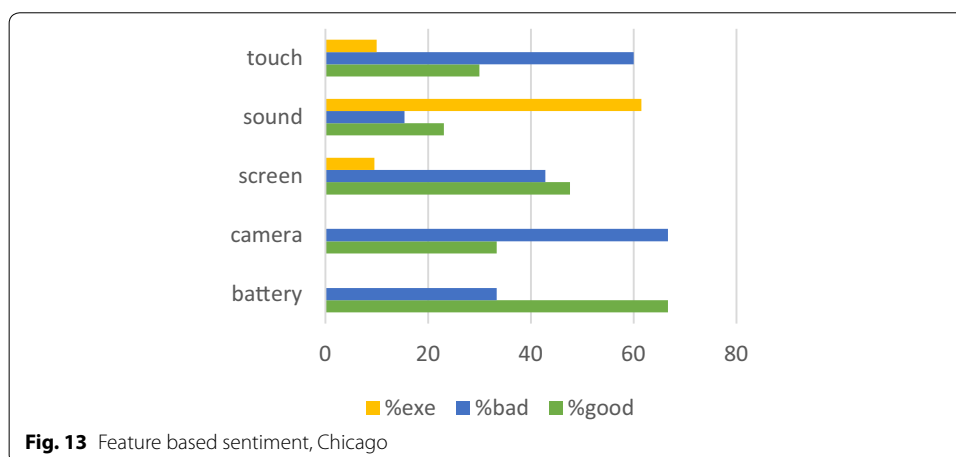
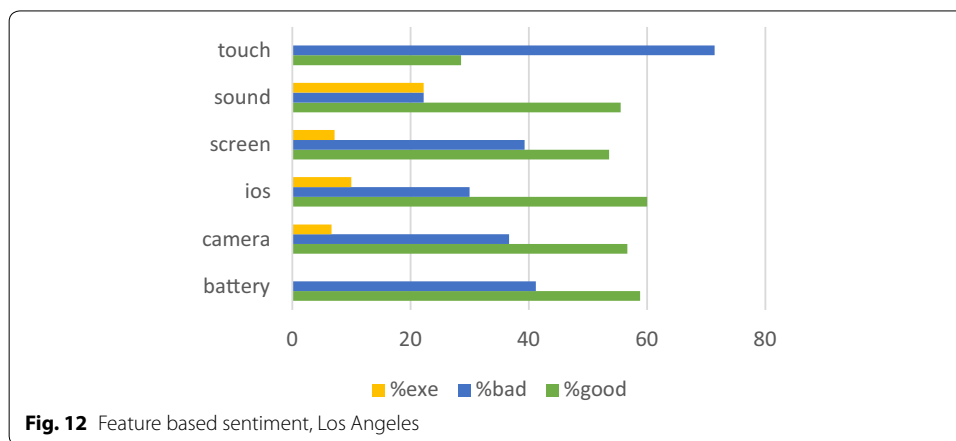
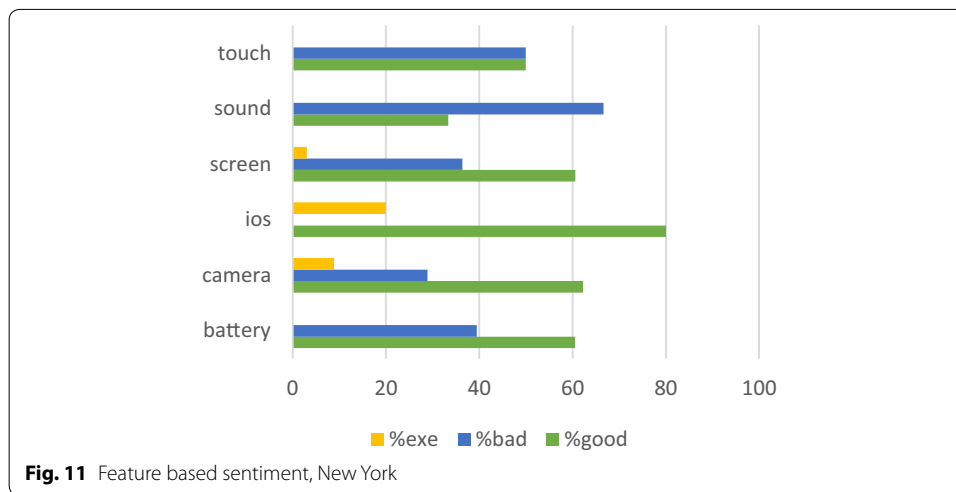
**Fig. 8** Average score per City



sentiments from both genders being present in the graph. This is usually a good indicator of diversity in these cities, meaning very mixed user opinions for the iPhone 6 are present. The similarities with previous comparisons are well highlighted. We got most excellent positive sentiment in Philadelphia and San Francisco with an average positive sentiment in Boston and Chicago. Here %bad, %good and %exe means percentage of tweets whose sentiment score was normalized and ranked as “bad”, “good” and “excellent” respectively.

Figures 10, 11, 12 and 13 below showcase feature specific sentiments for different cities. This assessment gives a very good idea about how popular the given features of the iPhone 6 are in different cities. This information is of crucial importance in marketing and advertisement concerns. For example, assume that a city shows high positive sentiments for the iPhone 6 device as a whole amongst all users. This means high level of popularity and greater sales. But this may not be the case. If the city in context has low positive or even negative sentiment towards a very important feature of the device, such as camera or screen, this could actually hamper sales. In our case, such a scenario is created in case of San Francisco. The overall sentiment for San Francisco is very high. But nearly 50 % users have regarded as the touch of the new iPhone as bad. This will adversely affect the final sales of the iPhone 6 because opinions and word of mouth are very important to people who are exploring blogs or tweets for a new smartphone to purchase.



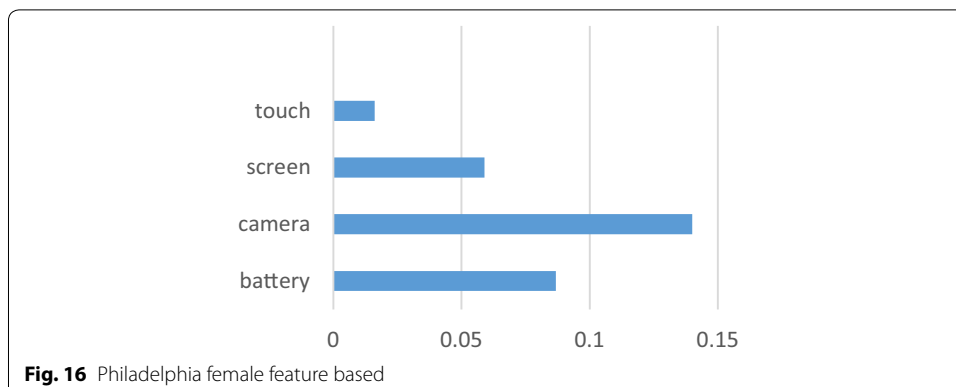
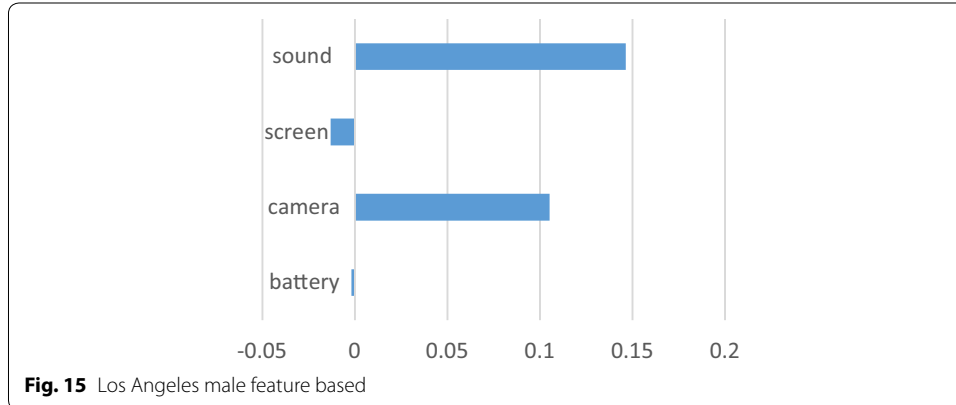
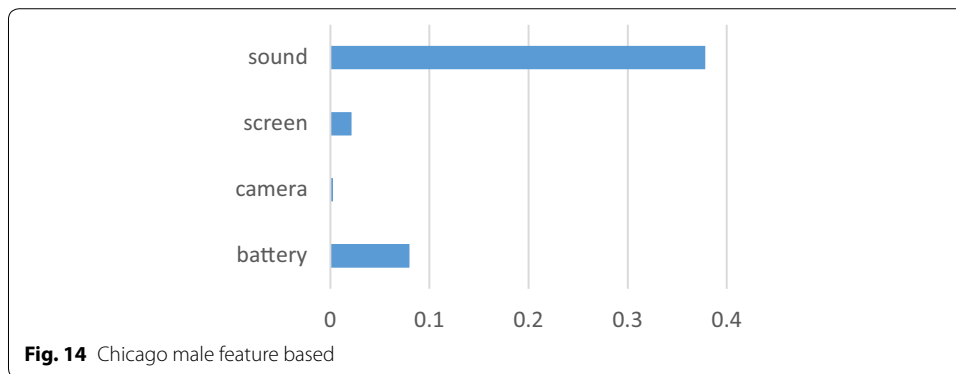


In some of the above charts we can see that one or two features are missing. This is because tweets related to that particular feature were not available in the given city. This is generally a result of complex querying and also because twitter API does not return



tweets more than 5 days old. Even though this is a good option, because there is a guarantee that the tweets are as recent as possible, but the problem is that more complex queries generally result in very few, sometimes even no tweets at all. Such is the case for the feature specific comparisons above.

Figures 14, 15 and 16 illustrate city and gender based average sentiment scores. This is the most in depth comparisons where all the variables involved have been taken into account. Generally lower battery and screen sentiments are again seen for Chicago and Los Angeles. The high positive sentiment for sound could be because Chicago is highly



populated by musicians and music schools as seen in the article by Renzulli (2015). These people showed praise for the sound quality of the iPhone 6.

Figure 16 shows the sentiment for female iPhone users from Philadelphia. This figure has a couple of signs uncommon to other comparisons. Here both the touch and screen features have received high positive sentiments. This is consistent with the fact that not all people faced the bending issue. Even though complaints did come up from majority of customers, significant fractions remained unaffected.

### Other applications

Even though a technological product has been studied in this research, it is worthwhile to mention at this point that the method described is also applicable to non-technical products, services or anything that has significant amount of exposure on social media. An example could be general elections. Nowadays all electoral candidates turn to social media for their campaign programs. So naturally it is expected that people will tweet about these politicians. So analyzing these tweets using the methods shown can give important insights about where the candidates stand in the elections. Keywords such as 'election', 'congress', and 'president' etc. or hashtags from the website (Hashtags for #election2016 in Instagram, Twitter, Facebook, Tumblr 2015), which contains popular hashtags related to the US election 2016, can be used to extract tweets. Generally it could be said that low sentiment tweets about a given candidate means lower popularity and general aversion amongst people, while high sentiment tweets will mean the opposite.

Another example could be determining public reactions to a change in, or introduction of new laws. Again the concept of the analysis would be similar, meaning that if people have positive mentions about the law and show their support via social media, this can be distinguished by a high sentiment score.

### Conclusion

In this research we discussed a methodology by which it is possible to determine the popularity/opinion/sentiment of a product in different locations across male and female users. For our analysis we chose the iPhone 6 as during the time of the research a reasonable amount of tweets based on the iPhone 6 was available. The number of tweets must be significant for accurate results. Therefore, even if a product does not have large number of tweets at any given moment, we could collect tweets over a period of several weeks or months or purchase large datasets from data centers. For the choice of location we select seven major cities in the United States. The reason behind this is also data availability. 35 % of all the worlds' tweets are from the USA with the remaining fraction heavily divided amongst all other countries.

But the methodology defined is much generalized and can be applied to tweets from any country for any product as long as a suitable number of tweets can be obtained.

Initially the tweets were filtered using Natural Language Processing tools. Only tweets which contained selected grammatical relations involving previously chosen keywords were selected. Each of the tweets in this filtered set were then given parts of speech tags. Each individual word in every tweet was assigned its own part of speech tag. After POS tagging the tweets were then processed using SentiWord which gave each tweet a

sentiment score. The program that we developed has been uploaded to GitHub, link is given in Ekram (2015), to aid future research on the topic.

The locations of the tweets were available from the data extraction phase because the twitter API takes a location parameter to find geo localized tweets.

The NamSor mining tool was used to classify genders for each of the tweets. This was required because twitter does not provide gender information. This was a very new approach because there is no previous sentiment analysis research where user names are used to determine the gender. The accuracy of the classification was also high mainly because of NamSor's statistical claim saying that over 96 % of the names from the United States are given correct gender classifications.

Finally the data was presented graphically and several comparisons to real world scenarios were made to justify the accuracy of the methodology.

The clearest deductions were the highly generalized negative sentiments towards the iPhone 6 screen and touch which is because of the bending issue plaguing iPhone 6 users since its release. And also a general high sentiment towards the iPhone 6 camera which has been praised by both general users and reviewers as seen in the review by Mansurov (2015) and photographers for having near lens based camera image quality and sharpness.

### **Limitations**

The twitter data that was collected was fairly good enough to demonstrate the usage of the method. The results were also very accurate when compared with real world scenarios. But nonetheless the amount of tweets collected was still very few. This happened because of complex querying and introduction of the location parameter. So a larger and richer set of tweets would generate even better results.

Even though NamSor is extremely accurate in gender classification there is still some error in that part. Ready availability of the gender with tweets would reduce that error even more.

Quality of tweets was also very low. Even after filtering there were quite a few tweets which resulted in a zero score from SentiWord, meaning that those tweets in no way express any sentiment towards the iPhone 6. Lower number of such tweets and larger number of expressive tweets would also yield more elaborate and diverse results.

### **Future work**

Further work on the methodology would involve even better NLP filtering with more grammatical relations being introduced. Also a custom SentiWord dictionary could be utilized which will contain scores for custom words which are not available in the stock SentiWord database. More advanced comparison approaches can be taken such as clustering etc.

A repetition of the experiment could be performed on a better set of tweets to obtain much better results.

### **Authors' contributions**

In this research we have discussed a methodology by which it is possible to determine the popularity/opinion/sentiment of a product in different locations among male and female users by analyzing tweets based on the product. For our analysis we have chosen one of the most popular products of Apple, e.g. iPhone 6. We have found mixed opinions most of which were consistent with general comments and opinions expressed by users about the new Apple product. All authors read and approved the final manuscript.

**Acknowledgements**

There is no acknowledgement from authors' side.

**Compliance with ethical guidelines****Competing interests**

The authors declare that they have no competing interests.

Received: 26 February 2015 Accepted: 5 August 2015

Published online: 22 October 2015

**References**

- Akhtar, N. (2014). Social Network Analysis Tools. In *Fourth International Conference on Communication Systems and Network Technologies* (pp 382–388).
- Beavis, G. (2015). iPhone 6 review. <http://www.techradar.com/reviews/phones/mobile-phones/iphone-6-1264565/review/1>. Retrieved July 16, 2015.
- Cho, S. W., Cha, M. S., Kim, S. Y., Song, J. C., Sohn, K.-A. (2014). Investigating Temporal and Spatial Trends of Brand Images using Twitter Opinion Mining. In *2014 International Conference on Information Science and Applications (ICISA)* (pp 1–4).
- Ekram, T. (2015). Tahmid140/twitter-opinion-mining. <https://github.com/tahmid140/twitter-opinion-mining>. Retrieved July 31, 2015.
- Hashtags for #election2016 in Instagram, Twitter, Facebook, Tumblr. (2015). <http://top-hashtags.com/hashtag/election2016/>. Retrieved July 30, 2015.
- Khanafarov, D., Luc, C., Wang T. (2014). Social Network Data Mining Using Natural Language Processing and Density Based Clustering. In *IEEE International Conference on Semantic Computing (ICSC)* (pp. 250–251).
- Kim, H.-G., Lee, S., Kyeong, S. (2013). Discovering Hot Topics using Twitter Streaming Data and Geographical Clustering. In *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp 1215–1220).
- Mansurov, N. (2015). iPhone 6 Plus Camera Review. <https://photographylife.com/reviews/iphone-6-plus-camera>. Retrieved July 21, 2015.
- Namsor. (2015). <https://github.com/namsor/namsor-api>. Retrieved July 30, 2015.
- Nithish, R., Sabarish, S., Abirami, A.M., Askarunisa, A., Navaneeth Kishen, M. (2013). An Ontology based Sentiment Analysis for mobile products using tweets. In *Fifth International Conference on Advanced Computing* (pp 242–247).
- Ostrowski, DA. (2012). Semantic Social Network Analysis for Trend Identification. In *IEEE Sixth International Conference on Semantic Computing* (pp 215–222).
- Ostrowski, DA. (2013). Semantic Filtering in Social Media for Trend Modeling. In *2013 IEEE Seventh International Conference on Semantic Computing* (pp 399–404).
- Renzulli, M. (2015). Top Music Cities in the USA. <http://usatravel.about.com/od/Top-Destinations/ss/Top-Music-Cities-In-The-Usa.htm#showall>. Retrieved July 27, 2015.
- Rubin, R. (2014). iPhone 6 bending: Common sense for an uncommon problem. <http://venturebeat.com/2014/09/28/iphone-6-bending-common-sense-for-an-uncommon-problem/>. Retrieved December 10, 2014.
- SentiWordNet. (2015). <http://sentiwordnet.isti.cnr.it/>. Retrieved July 30, 2015.
- Servia-Rodriguez, S., Fernandez-Vilas, A., Diaz-Redondo, R. P., Pazos-Arias, J. J. (2013) Comparing tag clustering algorithms for mining Twitter users' interests. In *2013 International Conference Social Computing (SocialCom)* (pp 679–684).
- SNLP Manual. (2015). Stanford Typed Dependencies Manual. [http://nlp.stanford.edu/software/dependencies\\_manual.pdf](http://nlp.stanford.edu/software/dependencies_manual.pdf).
- SNLP Group. (2015). The Stanford NLP Group. <http://nlp.stanford.edu/>. Retrieved October 4, 2014.
- Viklund, A. (2015). Free Map Tools. <http://www.freemaptools.com/>. Retrieved August 23, 2014.
- Williams, A. (2012). Abraham/twitteroauth. <https://github.com/abraham/twitteroauth>. Retrieved September 19, 2014.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---