Real-time Twitter Data Analysis of Saudi Telecom Companies for Enhanced Customer Relationship Management

Shaheen Khatoon

College of Computer Science and Information Technology Department of Information System
King Faisal University, Al Ahsa, Saudi Arabia

Abstract
In this research an automatic solution for helping telecommunication companies to gain better customer insight by utilizing real time Twitter data is proposed. To develop an enhanced Customer Relationship Management (CRM) in highly competitive market, it is very important for companies to understand how customers perceive and select specific services offer by them, compare to their competitors. While understanding the customer perception from conventional sources such as surveys, interviews and feedback is well studied and established, gaining insights from twitter is challenging due to several underlying issues in Twitter data, such as short message length, diverse colloquial linguistic patterns and representational richness. This study discusses the challenges of analyzing Twitter data and proposed a sentiment analysis approach to identify whether customers are enjoying with services or having bad experience. As a proof of concept, we present some preliminary results from telecommunication domain pertaining to three major Saudi companies, Mobily, Zain and Saudi Telecom Company (STC). To demonstrate the usefulness of work temporal and spatial changes towards sentiments of these companies are determined. This allow companies to monitor customer perception towards their own service and services provided by their competitors over time. Furthermore, by visualizing sentiment spatially in real-time, company can identify which region customer are having bad experience. This insight would allow companies to solve the problem in real-time fashion and provide better quality of service by redesigning product or service by considering the customer's needs.

Key words:
Sentiment analysis; Twitter; SentiWordNet; Natural language processing; CRM, Telecommunication

1. Introduction

With the advancement in information technology and fast communication, now a day it is much easier for customers to share their voice about product or service over Internet than ever before. This puts the customer into a position, where a simple tweet can tell hundreds or sometimes even thousands of other potential customers about their bad experience. This negative message might seriously damage the company reputation in highly competitive labor market. To avoid such risks an organization should have the ability to monitor their customer complaint automatically and act before it is too late. Hence, it is crucial for organizations to gain better customer insight utilizing various information sources, since, good customer's insight is the foundation of good Customer Relationship Management (CRM). Organizations can get insight about their customer by investing on developing conventual CRM systems, where they can maintain substantial information about their customers. By utilizing the information from their usage behavior, complaints and feedback companies can create different customer segments and offer services to individual satisfying their personal business needs. However, this requires IT to prepare data and modify data model to fit in individual's business scenario, which is a time-consuming process with delay response and offers a passive relationship of customer with company. As a result, companies are not achieving the results what they are expecting from their huge investment in CRM.

With the rise of social media such as Twitter and Facebook among many others, the customers are no longer limited to a passive role in their relationship with a company. They are receiving and generating more information about competitive products available anywhere in the market on their mobile devices. The chances of losing customers is higher if companies are not providing timely response and better services. To avoid such risks organization should be able to process customer opinion, product and services reviews, discover interesting patterns e.g. strength and weakness of services offered, level of acceptance of new service or product, how the industry is changing, what their competitor position in market from social media data. By utilizing data from such platforms combined with conventual CRM, it is believed the organization would able to better understand customer desires and experience which help them redesign product or services reflecting customer needs. In addition, conventional resources took long time in analysis process, now, the new data source has made it possible to execute in real-time fashion.

To do so there is a need of automatic solutions which monitor the customer activities on social media and generate alerts when critical condition occurs and display overall customer sentiment over time. This will allow top management to make important decision and solve customers' problems in real-time fashion. Many large organizations are already utilizing such systems, one example is the chain of Starwood Hotels and Resorts [1],

Manuscript received February 5, 2017
Manuscript revised February 20, 2017

where they utilizing the social media to stay connected with their customer, collecting the feedback, addressing the issues and helping them to make their travel decisions. However, in Saudi Arabia context generally and Saudi telecommunication specifically there is a minimum involvement of customers while designing product or services, and customers have to use those services due to monopoly of few companies in market. However, this case will not sustain longer, since, industry is changing rapidly and companies cannot monopolies any longer if they are not developing competitive intelligence by leveraging the hidden knowledge from contents generating by customers on social media.

By keeping the above issue in mind, this study is the initiative for Saudi companies towards development of competitive intelligence by identifying what opinion public is holding for them and their competitors. An automatic solution based on sentiment analysis is proposed to determine whether a tweet corresponding to these companies is a positive or negative. The level of positivity and negativity are called Sentiment polarity which is calculated by using SentWordNet (SWN) [2]. SWN is a WordNet (WN) based lexicon resource where each synset of WN has assigned three numeric values namely, positive, negative and objective. Each incoming tweet is matched with its corresponding synset in SWN and sentiment score is assigned to be classified as positive or negative. At any given time, company can view overall sentiment trend by aggregating sentiment scores across all tweets. Furthermore, managers can visualize which specific location customers have low sentiment. This would allow companies to solve the issues as well as redesign product or services by considering the customers need. To best of our knowledge this is the first study in Saudi telecommunication context and we believe the study will be helpful for these companies to tap into customer's insights to provide better service.

In the rest of this paper we present detail of the system that is developed for telecom companies' managers aiming to help them identifying what opinion customer are holding towards product and services offer by them and similar services offered by competitors. As an example, we present some preliminary results by capturing temporal and spatial changes in sentiments.

2. Related Work

Several tasks in social media involve the extraction and analysis of data relevant to specific subject of interest. For example Twitter data has been utilized in monitoring real-time occurrences of major events such as major sports events, outbreak of news stories, disease outbreak and the reactions of the users to events [3-7]. Twitter has also widely used as a source of information to predict financial market, movements in stock markets and other socioeconomic indicators [8]. Researcher has also utilized Twitter data to forecast the outcomes of elections through the analysis of tweets on the candidates and political issue [9, 10].

Majority of the studies in these areas of Twitter research employ sentiment analysis approaches to identify and evaluate the opinions of users expressed in their tweets. It is an active area of research, with the proliferation of blogs and social networks supported by advancement in Natural Language Processing (NLP) and text analytics techniques. It is aim to identify and evaluate the user opinions expressed in their text to identify and predict various business and social issues. Sentiment analysis is widely used in domain of product review by utilizing data from online review sites to identify positive or negative feedback about product such as smart phone. In film review it is used to give a rank to specific film with a feedback range from one to five stars[7, 11, 12].

Recently researchers has utilized the Twitter data to determine the consumers sentiment towards a brand to identify emerging issues related to brand which allow brand managers to make better decisions [13, 14]. For example Cho et al. [15], determined the spatial and temporal changes of different brands from Korean tweets. They developed a Korean polarity dictionary to compute sentiment polarity, whereas, sentiment classification is done by using SVM and Naïve Bayes to automatically analyze the sentiment of each tweet. Similarly, Scheuer et al. [16], used supervised learning to extract brand perception from twitter. They developed a classifier by utilizing labeled data with predefined concept in Retailer domain. By evaluating the same classified across different domains, they found different level of complexity in classifying the pre-defined concept.

In above studies two main approaches for sentiment analysis are used, namely, Machine learning and lexicon based approaches. Machine learning [17] is a supervised or semi supervised approach which requires pre-classified dataset on specific domain called training dataset to predict sentiment label of new incoming data. Some well-known classification algorithms such as k-Nearest Neighbors, Naïve Bayes and Support Vector Machines, etc. are widely used to predict sentiment class. Whereas, lexicon based approach [12] uses the lexicon resource of word from predefined list or dictionary, where a positive or negative score is assigned to each word. Sentiment polarity of unseen document is then calculated by identifying its semantic orientation in predefined lexicon list using different text processing techniques. On similar line of research, we used lexicon based approach, where SWN is used as lexicon resource to compute sentiment score of tweet collected from Twitter in telecom domain.

3. Proposed Approach

In this section, we will discuss our proposed approach to assess the sentiment pertaining to three major Saudi telecom companies. The proposed approach receives data from twitter and calculate sentiment scores using SentiWordNet (SWN) version 3.0 [2], for each tweet and categorized them as positive or negative. The abstract view of proposed approach is shown in Figure 1. It consists of four main modules as described below:

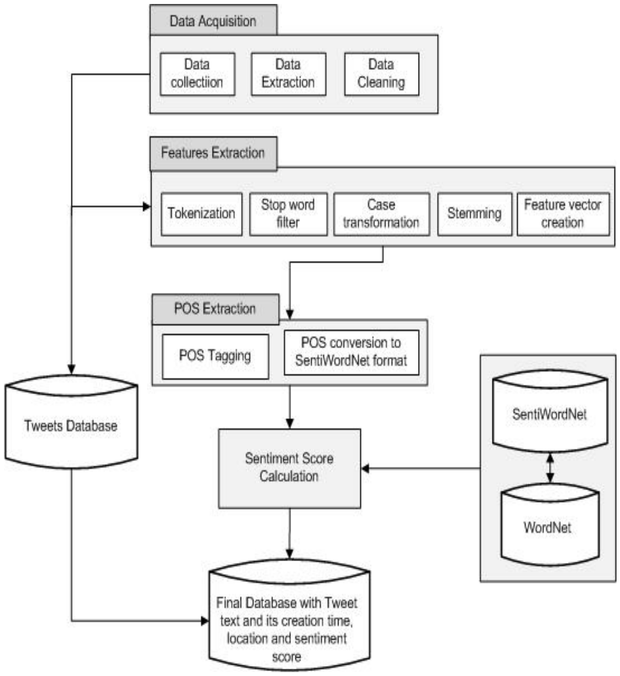

Fig 1. Abstract View of proposed approach.

3.1 Data Acquisition

For the purpose of this study we first need to extract data from Twitter. To do so data acquisition module is developed consisting of data collection, data extraction and data cleaning procedures. The overall process of data acquisition is shown in Figure 2.

Data Collection

To achieve our task of getting customer insight for telecom domain, we selected three Saudi telecom companies Saudi Telecom Company (STC), Mobily and Zain to collect data. Twitter search APIs have been used to retrieve 1000 tweets from each company using filtering terms that are relevant to their official and their service accounts (@STC #STC, @STCcare, @ZainHelpSA, #Zain, @Zainksa, #Mobily, @Mobily1100 and @Mobily) defined by longitude and latitude of Saudi Arabia.

Data Extraction

Data retrieved from twitter contain several attributes such as user names, user Id, and source, retweet count, tweet text, time, geolocation and tweet id. This study only requires tweet text, created at and geolocation, where, tweet text is used to calculate sentiment, time and location information are used to identify spatial and temporal changes in sentiment. Geolocation is obtained by providing values to 'location' parameter as latitude, longitude and radius information in query. For each tweet its text, 'created at' and 'location' is extracted and stored in MySql database for further processing.

Data Cleaning

After data extraction, data cleaning and normalization has been performed. At first, specific twitter noise, such as urls, numbers, punctuations, special characters and human expression are removed. To make the dataset ready for applying sentiment it is further cleaned by correcting typos and slangs. Regular expression and simple rules are used to perform data cleaning and normalization operations.

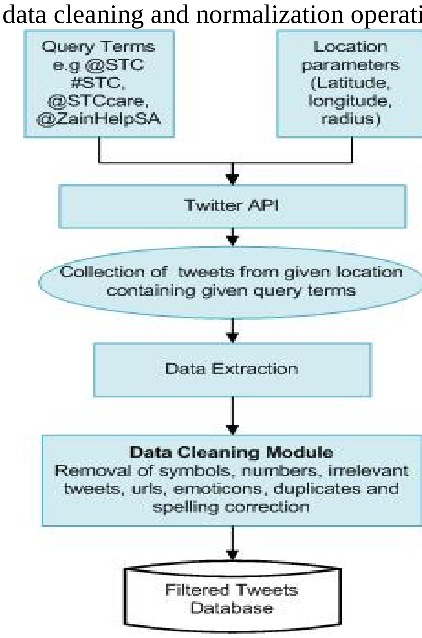

Fig 2. Data acquisition procedure

3.2 Feature Set Extraction

All classifier work with feature set, which is a simple dictionaries mapping feature name to feature value. In order to do so bag of word model is used to generate feature vector of each tweet, where occurrence of each word is used as feature and Term Frequency Inverse Document Frequency ($tf.idf$) is used as feature value. To generate feature vector from tweet several basic Information Retrieval (IR) techniques are applied described as following:

Tokenization

It is the process of splitting the input text into sequence of individual units called tokens. Tokenization can be applied at different level such as paragraph, sentence and word level. Since, we need words as feature vector therefore word level tokenization is performed using non-letter character such as "punctuation marks" and "white spaces" as a separator, thus resulting tokens contain only letters.

Stop Word removal and lower case transformation

Once tokens are generated, the letters in each token is reduced to lower case to map same tokens with different case into identical token. In next step stop words are filtered from tweets by removing every token which equals a stop word from built-in English stop word list. To further reduce corpus size query terms such as STC, Mobily and Zain, which were appearing in every token are also removed.

Stemming

It is the technique used to map different grammatical forms of word to a common term, e.g. the words 'writing', 'wrote' and 'written' all are mapped to word 'write'. The different forms of verb and ending with singular and plural does not matter in task of analysis, since we are concerned about context of tweet. In this work dictionary based stemmer is used which reduce terms to base form using Wordnet dictionary which lists all words together with their stems. The stemmer take tokenized words as input and generate list of stem.

Document vector creation

After tokenization, has been performed, the word vector representation of all tweets is generated using vector space model also known as bag of word model. In this model a document is represented by vector that denote the relevance of given set of terms (words) in the document. The relevance of terms i in document j denoted as v_{ij} is calculated by using $tf.idf$ scheme with:

$$v_{ij} = \frac{f_{ij}}{fd_j} \log\left(\frac{|N|}{ft_i}\right)$$

Where, $|N|$ = total number of documents, f_{ij} = total number of term i in document$i$, fd_j = total number of terms in document j, ft_i= total number of documents in which term i occur at least once. The resultant vector for each document is normalized by using Euclidian unit length. In this way we can represent entire set of tweets as document collection and the document frequency for the term is the number of relevant document in which term appears. Doing this we can rank positive and negative tweets by identifying terms that most characterize the sentiment orientation in given tweet.

An example of partial word vector dataset is shown in Table 1, where each column shows normalized $tf.idf$ value of given word in document and each row representing document in collection. The number of columns in a word vector is a function of the number of unique terms in the document collection.

Table 1. Bag of word model Example

	enjoy	**claim**	**block**	**…**
D1	0.72	0.0	0.0	…
D2	0.0	0.54	0.57	…
…	…	…	…	…

A very high dimensional word vector spaces with several thousand attributes can easily be generated with large number of documents in collection. To mitigate with high dimensionality of vector space model, stop word removal and stemming has already performed to reduce the final number of terms.

3.3. Part of Speech Extraction

After applying basic IR techniques to generate a feature word vector, Part of Speech (POS) information of each feature is extracted. POS information important to detect sentiment polarity of a word as the sentiment of word may vary with POS tag. For example, novel is scored positively when used as adjective, whereas categorized as neutral when used as noun.

Part of Speech Tagging

To assign POS tag to every word in tweet, Stanford POS tagging utility is used. To do so each tweet is provided as an input to this module which assigned tag to each word. For example the tweet "disappointed from the poor service received" is passed to this module which is tagged as "disappointed |JJ from|IN the| DT poor |JJ service |NN received |VBD".

Part of speech Mapping

By recognizing the fact that SWN only four accept basic POS (nouns, adjectives, adverbs and verbs) to calculate polarity of word, we have developed POS mapping (POS mapper) module. The mapper take tagged words generated by POS tagger and normalized it to SWN format. For example, the tag generated for "disappointed" is "jj" which is mapped to "a" as adjective. Similarly, "NN" for service is normalized to 'n' as noun. Any other tag which does not contain adjective, noun, verb or adverb is discarded since it does not contribute towards sentiment polarity.

3.4 Sentiment Scoring

This module deal with the assignment of sentiment scores to each tweet using SWN (Version 3.0) [2]. SWN is a sentiment analysis lexical resource made up of synset from WN, where all WN synsets are annotated for degree of positivity, negativity and neutrality/objectivity along with POS information. Each synset is a group of terms that are synonyms of one another, where it is assumed that different senses of the same term may have different opinion-related properties. For example, the synset [estimable #3] corresponding to the sense "may be computed or estimated" has an objective score of 1, positive and negative score of 0, while the synset [estimable#1] corresponding to the sense "deserving of respect or high regard" has a positive score of 0.75 and 0 objective and negative score. Each of the three scores ranges from 0.0 to 1.0, and their sum is 1.0 for each synset. A sample from SWN is shown in Table 2, where each row is defined in the form $\langle POS, SID, sen^{+}, sen^{-}, ST, G\rangle$, where POS is part of

speech, since terms are categorized into parts of speech derived from WordNet (a, v, n and r correspond to adjective, verb, noun and adverb), SID is SWN key which maps the synset to the WN, sen^+, sen^- are the positive and negative scores of term T_i. $ST[vi] = \{s0, s1, s2, \ldots sn\}$ are the synset of vi and G is gloss description of T_i, which contain the meaning and sample usage of the terms present in the synset. A synset is "objective" if $1 - (sen^+ + sen^-) = 1$. Also, sum of $(sen^+ + sen^- + sen^-) = 1$.
To calculate sentiment of a tweet using SWN each word in tweet is tagged with corresponding part of speech using POS tagger. Since SWN only recognizes nouns, adjectives, adverbs and verbs, any parts of speech other than these four has to be map to any of these. Therefore, POS mapper is developed which normalize POS to these four available.

Table 2. Sample SentiWordNet entries

POS	SID	Sen+	Sen-	ST	Gloss
a	02113449	0.5	0.125	unsanitary#1 unhealthful#1 insanitary#1	not sanitary or healthful; "unsanitary open sewers"; "grim and unsanitary conditions"
v	02110793	0.5	0	enjoy#4	have for one's benefit; "The industry enjoyed a boom"
n	14259320	0	0.875	ophthalmitis#1 ophthalmia#1	severe conjunctivitis
r	00275035	0	0	asleep#1	into a sleeping state; "he fell asleep"

Using combination of part of speech and word itself, SWN assigned a numeric score to given word between -1.0 and 1.0, where -1.0 means very negative and 1.0 means very positive. As a tweet contain few words, it is rather convenient to compute score for each word. Since, one word may have multiple synset (senses) based on its part of speech, therefore, score for each sense is computed and total score of a term is calculated as the average value of all terms found.

$$Score(Term_i) = \frac{1}{TotalSyn} \sum_{i}^{n} Score(i)$$

Where $Score(Term_i)$average is score of given term, and $TotalSyn$ is the total number of synset of all possible POS of term T_i found, $Score(i)$ is the score for each synset. Sentiment score for individual term is calculated as the average value of all synset found for term T_i. Finally, sentiment of each tweet is calculated by taking the average value of all terms found in tweet. A complete example of giving score to tweet "disappointed from the poor service received" is shown in Table 3, where, first column contains each term in given tweet, 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 using mapping function. The fourth column is the SWN score. The total score of a tweet is the average of all the individual word scores.

Sentiment scoring example

Term	POS Tag	Normalized POS	Score
disappointed	JJ	a	-0.5
poor	JJ	a	-0.875
service	NN	n	0
received	VBD	v	0
Total score			**-0.344**

4. Results Analysis and Visualization

The proposed approach is analyzed using the three Saudi telecom companies STC, Mobily and Zain. Data is collected over a period of two weeks. Data preprocessing is done by using different regular expression and simple rules using Java. Tokenization, stemming, stop word removal and POS tagging is done by using NLTK[18]. Sentiment scores are calculated using SWN [2]. Once the tweets are processed and scored using aforementioned steps, the resultant dataset along with temporal and spatial information for each of the telecom company is stored in MySql database for subsequent analysis and visualization.
To demonstrate the usefulness of this work and to illustrate the sentiment trends we have constructed a sentiment time series pertaining to particular telecom company. The sentiment time series depicted in Fig. 3 shows nationwide daily sentiment trend over a period of 15 days, where each data point is aggregated across all tweets on given day. The x-axis depicts the given time while the y-axis shows the sentiment index ranging from -1 (extreme negative) to 1 (extreme positive). For the sake of privacy name of three major telecom companies are replaced with TCA, TCB and TCC. The figure shows sentiment trends of TCA has extreme sentiment polarity, whereas TCB has extremely negative polarity only once in duration of 15 days and TCC has comparatively positive sentiment. This analysis is very useful for companies' manger to investigate what went wrong on specific day to provide better quality of service based on customer's perception.

Fig 3. Daily sentiment: each datapoint is the average of daily sentiment score for a company over the priod of 15 days

Beside nationwide sentiment score, by utilizing this research, companies' managers can zoom-in to understand the perception with respect to different aspect. For example, manager, might interested to visualize the customer perception in specific region. To do this manager can select the location from drop down list from user interface intended to develop for companies. System will display the average sentiment score for selected location by summing up scores for all the tweets across given location. Location based sentiment is calculated using following equation:

$$SentScore(loc_j) = \frac{1}{N}\sum_{i=1}^{n} SentiScore_i$$

Where, $SentiScore_i$ is sentiment score for each tweet, loc_j is location selected and N is the total number of tweets in selected location. This assessment give a very good idea about how sentiment varies across region and which particular region has low sentiment. This information is crucial for a company for advertisement and marketing point of view. For example, if sentiment is low for a particular region company can offer low cost packages or other kind of personalized packages to build its reputation among users. Fig. 4 shows location based sentiment across four main cities of Saudi Arabia for each telecom company, where vertical axis shows sentiment index and horizontal axis shows city. It can be observed that TCA is liked more by Riyadh based customers and equally balanced among Damam and Madina customers, however, it is least liked by Jeddah based users. TCB is more trusted by Jeddah, equally balanced among users of Riyadh and Madina and least trusted by Damam users. Similarly, TCC is more liked by Dammam users and least liked by Madina users. Overall city based statistics shows TCC has comparatively high location based sentiment during this study.

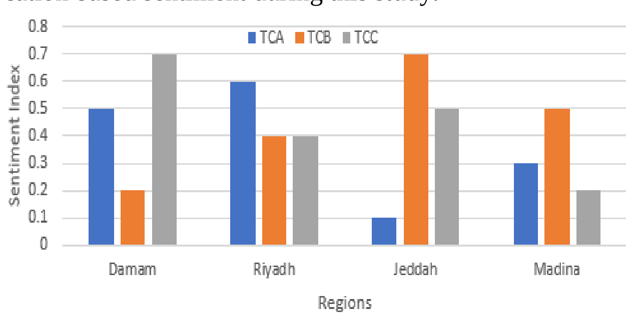

Fig 4. City based sentment where each data point is average of total tweets in given region

Similarly, company managers can compare these spatial and temporal changes with their competitor's perceptions. It will greatly help the companies to offer services that reflect the customer interests. The study is intended to develop a graphical user interface to be provided to company's manager to support data exploration through free text and faceted search.

5. Discussion and Future Work

This study used lexicon based approach for sentiment calculation utilizing SWN. However, there may be chances where some word polarity is not found in SWN. Currently, those cases are not handled, since this study present the preliminary result from telecom domain towards development of competitive business intelligence system. However, the words which are not found in SWN, are saved in separate list to construct domain specific dictionary in next phase.

Furthermore, in specific domain like telecom, most of the words have one sentiment class in SWN since it used word level information with its POS to calculate score, whereas their occurrence in the annotated dataset shows strong inclination with the other sentiment class. For example, consider tweet "STC makes their competitors look stupid for not considering of service 'X' first". SNW has classified this tweet as negative by assigning negative sore of 0.75 to word "stupid". Although this tweet should be considered as positive. It shows that simple lexical based approach using bag of word model is not good enough to capture correct sentiment, since sentiment is greatly affected by the differences of the characteristics of texts by context and domain. To address this limitation, in future research word scores will be computed by using supervised learning with annotated texts. The pre-classified data ('training data') will help in determining whether a text is having a positive or negative sentiment by identifying classification of similar text in pre-classified dataset. The lexical based approach combined with supervised learning will significantly improve the accuracy of proposed system.

It has also been observed during this study that tweets tends to be predominantly neutral, with far fewer expressing positive or negative sentiment. This contrasts with certain sentiment analysis domains e.g. product reviews, which tend to be predominantly positive or negative from relatively small number of reviews. It poses a great challenge in data collection and filtering process. It is assumed that with large data sample and domain specific lexicon result will be more consistent with actual perception of users. This study present preliminary results on small dataset, however, more accurate sentiment analysis by using large dataset combined with domain specific is a part of future research

6. Conclusion

In this research an automatic solution to determine customer perception towards Saudi telecom companies is discussed. A reasonable number of tweets are collected by using query terms pertaining to official twitter account of these companies along with location parameters. Tweets are filtered and cleaned by using NLP tools and only those

tweets are stored in final database which contains sentiment oriented words. After applying some basic IR techniques, POS tagging is done, where each individual word in tweet is tagged with it POS information. Tagged words are then processed by using SNW, where each word has given a sentiment score. Finally, score for each tweet is computed by taking the average value of all words in tweet.
To illustrate the usefulness of proposed approach to potential telecom companies, daily sentiment time series is constructed which will help potential company's manger to monitor change in sentiment trends over time. Furthermore, location based visualization of sentiment is also constructed to identify the spatial changes in sentiment. We believe, this work will help telecom companies to build better customer relationship combined with traditional CRM

References

[1] C. Muller, Impact of the Internet and Social Media on the Hotel Industry: Grin Verlag Ohg, 2011.
[2] S. Baccianella, et al., "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," in LREC, 2010, pp. 2200-2204.
[3] S. Khatoon, "Multi Agent Distributed Clustering Framework for Large Scale Social Data: Twitter Case Study," International Journal of Computer Science and Information Security, vol. 14, pp. 226-236, 2016.
[4] M. Walther and M. Kaisser, "Geo-spatial event detection in the twitter stream," in European Conference on Information Retrieval, 2013, pp. 356-367.
[5] D. Corney, et al., "Spot the ball: Detecting sports events on Twitter," in European Conference on Information Retrieval, 2014, pp. 449-454.
[6] E. Aramaki, et al., "Twitter catches the flu: detecting influenza epidemics using Twitter," in Proceedings of the conference on empirical methods in natural language processing, 2011, pp. 1568-1576.
[7] R. Nithish, et al., "An Ontology based Sentiment Analysis for mobile products using tweets," in Advanced Computing (ICoAC), 2013 Fifth International Conference on, 2013, pp. 342-347.
[8] J. Bollen, et al., "Twitter mood predicts the stock market," Journal of computational science, vol. 2, pp. 1-8, 2011.
[9] A. Bakliwal, et al., "Sentiment analysis of political tweets: Towards an accurate classifier," 2013.
[10] A. Bermingham and A. F. Smeaton, "On using Twitter to monitor political sentiment and predict election results," 2011.
[11] W. Chamlertwat, et al., "Discovering Consumer Insight from Twitter via Sentiment Analysis," J. UCS, vol. 18, pp. 973-992, 2012.
[12] S. A. A. Hridoy, et al., "Localized twitter opinion mining using sentiment analysis," Decision Analytics, vol. 2, p. 8, 2015.
[13] M. Ghiassi, et al., "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network," Expert Systems with applications, vol. 40, pp. 6266-6282, 2013.
[14] W. He, et al., "Social media competitive analysis and text mining: A case study in the pizza industry," International Journal of Information Management, vol. 33, pp. 464-472, 2013.
[15] S. W. Cho, et al., "Investigating temporal and spatial trends of brand images using twitter opinion mining," in Information Science and Applications (ICISA), 2014 International Conference on, 2014, pp. 1-4.
[16] M. Shmueli-Scheuer, et al., "Greaaaat bargains starting from just 99p!!!!:-): Brand perception in the social media," in Proceedings of the 1st international workshop on Multimodal crowd sensing, 2012, pp. 19-20.
[17] B. Pang, et al., "Thumbs up?: sentiment classification using machine learning techniques," in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, 2002, pp. 79-86.
[18] S. Bird, "NLTK: the natural language toolkit," in Proceedings of the COLING/ACL on Interactive presentation sessions, 2006, pp. 69-72.