Global Journal of Enterprise Information System

DOI: 10.18311/gjeis/2017/15616

Location based Twitter Opinion Mining using Common-Sense Information

Amita Jain[1*] and Minni Jain[2]

[1]Department of Computer Science and Engineering, Ambedkar Institute of Advanced Communication Technologies and Research, Delhi, India; amita_jain_17@yahoo.com

[2]Department of Computer Science and Engineering, Delhi Technological University, Delhi, India

Abstract

Sentiment analysis research of public information from social networking sites has been increasing immensely in recent years. Data available at social networking sites is one of the most effective and accurate source to identify the public sentiment of any product/service. In this paper, we propose a novel localized opinion mining model based on common sense information extracted from ConceptNet ontology. The proposed methodology allows interpretation and utilization of data extracted from social media site "Twitter" to identify public opinions. This paper includes location specific, male- female specific and concept specific popularities of product. All extracted concepts are used to calculate senti_score and to build a machine learning model that classifies the user opinions as positive or negative.

Keywords: ConceptNet, Natural Language Processing, Sentiment Analysis, SentiWordNet

Paper Code: 15616; **Originality Test Ratio:** 08%; **Submission Online:** 11-Jan-2017; **Manuscript Accepted:** 15-Feb-2017; **Originality Check:** 18-Feb-2017; **Peer Reviewers Comment:** 25-Feb-2017; **Double Blind Reviewers Comment:** 12-April-2017; **Author Revert:** 03-May-2017; **Camera-Ready-Copy:** 10-May-2017

1. Introduction

Sentiment analysis of data collected from social networking sites is an effective mean of discovering public opinion. Today, many companies conduct paper or online surveys to collect customer opinions about their product/service. In paper based surveys very less customer base can be involved and there is no assurance of honest and complete opinions. Therefore paper based surveys are not an effective approach to collect product opinions from public. With the emergence of social networking applications and sites, people tend to express their opinion about the product on their twitter or Facebook profiles. Social media sites have a lot of data related to people's opinions for services/products they use. Hence mining opinions from social media sites is a much more advanced method for market analysis. A lot of work has been done on opinion mining from twitter, most of which focuses on people's sentiment towards various topics. But analyzing twitter data in this manner gives a much generalized idea. To make it more specific, opinion mining can be performed on social media data from explicit locations. Our approach is to find the product sentiments in specific locations. To determine the popularity of a given product in several locations, a large data set of Twitter is analyzed. Tweets are a reliable source of information mainly because people tweet about everything and anything they do including purchasing new products and reviewing them. This will allow companies to emphasis their marketing expenditures on areas where sentiment is low, while keeping minimum advertisement in areas of high popularity locations. This proposed research work deals with outcome prediction and explores localized outcomes based on common sense knowledge.

Common-sense knowledge represents basic understanding that people gain through experiences[1]. Commonsense, is required to properly identify sentiments in natural language text, for instance, the concept "small queue" is positive when referring to a post office, while the concept "small room" should be understood as negative in a hotel review. The concept-level sentiment analysis outperforms other existing research work because it preserves the semantic association with multiword expressions. For commonsense knowledge, in this paper ConceptNet ontology is used. An AI tool "NamSor" is used to exactly deduce gender from a person's user name. After completion of the analysis phase, experimental results are presented.

The rest of the paper is organized as follows. Section 2 describes prior research work that has been done to determine sentiment of twitter data. Section 3 discusses the entire procedure of sentiment analysis including data extraction & preprocessing, feature extraction and sentiscore calculation. The results of the experiment have been visualized graphically and also interpreted in section 4. Finally in section 5, discusses conclusion and future work respectively.

*Author for correspondence

2. Related Work

Various sentiment analysis methods have been proposed in literature to classify user opinions. However, most previous existing methods mainly rely on simple keywords spotting, syntactic information and POS tags. Mullen and Collier[2] proposed a method to expand the concept set based on Osgood's theory of semantic orientation. Turney's[3] described a semantic orientation method for supervised sentiment classification. Gelfand et al.[4] have developed a method based on sematic relation graph to extract concepts from a documents. They used the relationships between words extracted from a lexical database to form concepts. Agarwal et al.[5] proposed sentiment analysis with dependency- based semantic parsing. Hriday et al.[6] developed a location based methodology to identify sentiment of opinions from twitter. Nithish et al.[7] mainly focused on market reaction of smart phones using sentiment analysis. They intended to determine the explanation of factors that influences a product rating. DBSCAN: A clustering based data mining algorithm is proposed by Khanaferov et al.[8] to analyses tweets from same domain.

3. Proposed Methodology

Proposed approach for concept-level localized twitter opinion mining comprises three main steps 1) Data extraction and preprocessing 2) Concept extraction and 3) sentiment score calculation.

3.1 Data Extraction and Preprocessing

In the proposed work, tweets from twitter are used as a source of data. To extract an extensive amount of Tweets a public API of twitter is used. Each extracted tweet has several kind of information like tweet_id, text, username etc. Out of these extracted information only the tweet_id and text is useful for this work.

For this research, tweets are extracted from six major cities in the India. Due to the language constraints and data availability, the choice of cities is very limited. Six major cities selected for tweets collection and experiments are: New Delhi, Chennai, Mumbai, Bangalore, Pune, Surat. The longitude and latitude of each major city is used to define the city. Free map tools by Viklund are used to select estimated radius of coverage for each city[9]. The values of radius, latitude and longitude are assigned to the parameter locations in the query build for API.

For implementation and testing product iPhone7 is selected. Using the proposed approach it is possible to evaluate popularity of any product but with a condition that the product must have a good amount of data available on twitter. At the period of the research a large amount of tweets related to iPhone7 were available. So the tweets that contain the word "iPhone 7" are extracted. The most or least popular concepts of the iPhone7 are determined using some keywords to collect concept specific tweets. A case is 'iPhone7 battery'. This query will return all tweets containing the terms, iPhone 7 and battery together. Other keywords used are jet-black, 3.5 mm jack, camera, iOS, iTunes, screen, sound, and touch. From each the tweet text, location, and username are extracted.

3.2 Concept Extraction

In order to extract useful concepts, the Stanford Natural Language Processing tool by The Stanford NLP Group (SNLP Group 2015) is used[10,11]. This SNLP tool provides outputs that are grammatical relations between terms in a sentence. In this proposed work, to extract concepts from tweets, three main relations (also called dependencies) are used. These relations are **nsubj, dobj, amod**.

- The first relation is **nsubj**. "This relation is used to find dependencies between nouns and adjectives or verbs which are complementing the noun in a sentence". This relation provides an important knowledge that whether a sentence is complementing a noun or not. An instance of this **nsubj** relation is:

"My iPhone7 camera is amazing!" For this sentence, the concept **(camera, amazing)** is extracted.

- The second relation is **dobj**. "It is a direct object relation and used to find direct objects that a verb is referring to in a sentence". An example would be

"Hate the sound of iPhone7!" For this tweet the system extracts the concept **(Hate, sound)** and also **(Hate, iPhone7)**.

- The final relation is **amod**. "It is adjectival modifier relation and used to identify any adjectives used in a sentence to modify a noun phrase". For instance.

"Got my new Jet-black iPhone7, feeling wow!!" For this tweet, concept **(iPhone7, jet-black)** is extracted

3.3 Extracting Concepts from tweets using Common-Sense Knowledge

To extract more concepts related to product using common-sense knowledge 'ConceptNet' is used. "ConceptNet is a large semantic network consisting of large number of common-sense concepts" [12]. To mine several inferences from the text, commonsense knowledge available in ConceptNet can be used. It comprises vertices 'concepts' linked by edges 'relations between concepts'. Some of available relationships in the ConceptNet are: MadeOf, IsA, AtLocation, EffectOf, DesireOf, CapableOf etc[12]. Now to extend concept list, extracted concepts from previous steps using three dependencies relations are sent as query to ConceptNet. For instance, let the concept 'birthday party' is sent to ConceptNet as query, then output would be concepts such as buy gifts, cake. In ConceptNet, relations related to concepts are :"cake – AtLocation-- birthday party" and "buy gifts

– UsedFor—birthday party". These new set of concepts provide more knowledge. This combination of dependencies with commonsense knowledge offers a better understanding of text to the system[13]. The proposed system allows the machine to better understand the matter and the meaning conveyed by the natural language text.

3.4 Calculation of Sentiment Score of Tweets

To calculate the sentiment of extracted tweets a numeric metric is required. SentiWordNet (2015), is used to calculate overall sentiment of tweets[14]. SentiWordNet provides a score lies between −1 to 1, where higher value indicates the more positive sentiment and lower value indicates to lower sentiment. Score from SentiWordNet for each word in a tweet (containing at least one concept extracted on previous section) is obtained and then addition of all sentiment values to get a score for whole tweet. There is an important thing that SentiWordNet takes terms and their part of speech as input. The part of speech of the terms depends entirely on the sentence itself. So to map each term of sentence to its part of speech, a POS tagger is used.

The sentiment for each term is obtained using SentiWordNet and then the term scores are added to get final sentiment score of tweet, Eq. (1).

$$Score\ (location)_j = \frac{\sum_{i=1}^{n} Senti_{Score_i}}{n} \quad (1)$$

where, n is total number of tweets, Senti_Score i is SentiWordNet score for each tweet, Location j is refers to one particular city.

The scores calculated in this method do not follow any scale so it is necessary to normalize these final scores to get fixed sentiment scores for the tweets. To get information about users' gender from the tweets, a tool called "Nam-Sor" (2015) is used. It is an existing data mining tool[15]. The entire method is illustrated in Fig. 1.

4. Results and Discussion

Several comparisons are made to properly understand the variations in sentiments. Three comparisons are done to demonstrate the sentiment trends using proposed approach. These are:

1. *Countrywide Sentiment as Percentage:* Overall national sentiment of product.

2. *Countrywide concept Average Score:* This score includes concepts related to product for all the six cities. This score provides the overall view of sentiment towards the iPhone7 features.

3. *Gender Concept Average per city:* In this sentiment scores, all six cities are grouped on gender (male/female) basis for distinct concepts. It comprises all the mentioned variables that are gender, product features and specific location.

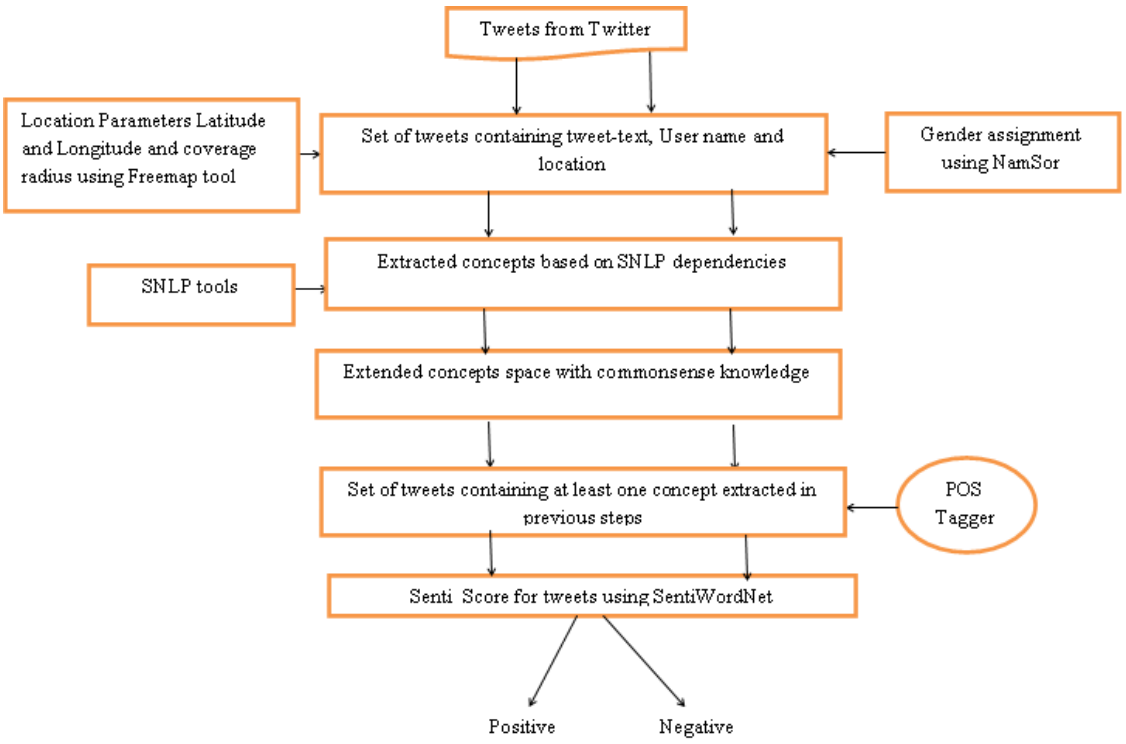

Figure 1. Flow diagram of proposed approach to location based sentiment analysis.

Originally 1040 tweets are collected using the twitter API. But after filtering only 552 are left for the further process. The tweets other than 552 are not valuable in for opinion mining. In this work, product search parameter and concept search parameter and location parameter are used so the count of tweets that obtained is lower than expectations.

For better comparability and easy understanding, all of the comparisons are explained using graphs. As shown in Fig. 2, 75% above users believed that iPhone7 is a worthy product. Usually excellent sentiments are more challenging to find because this requires huge number of tweets which cover words with very high positive SentiWordNet scores.

Next, the popularity of product iphone7 on the basis of its features is shown in Fig. 3. As per the results, feature iOS and Camera of the iPhone7 got highest positive sentiments. Result shows low positive sentiment of users about Jet-black color and battery of iPhone7. Now the final graph shown in fig. 4 includes results based on location, and gender (male/female). There are consistent sentiments for the both female and male with the touch and screen features. Observing Fig.4, the cities, New Delhi and Pune shows most positive sentiment scores with Mumbai following on close by cities Bangalore and Surat have moderate level in positive sentiments while the city Chennai has lower positive sentiments.

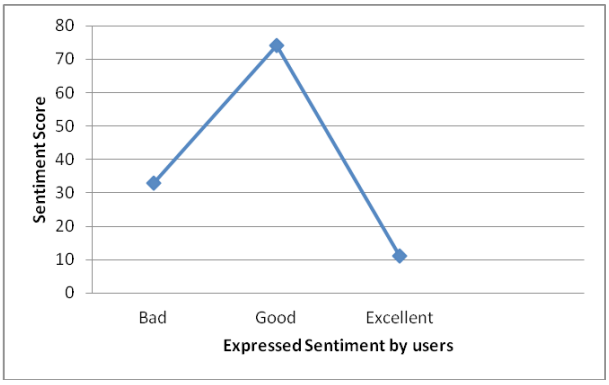

Figure 2. Overall countrywide sentiments as percentage.

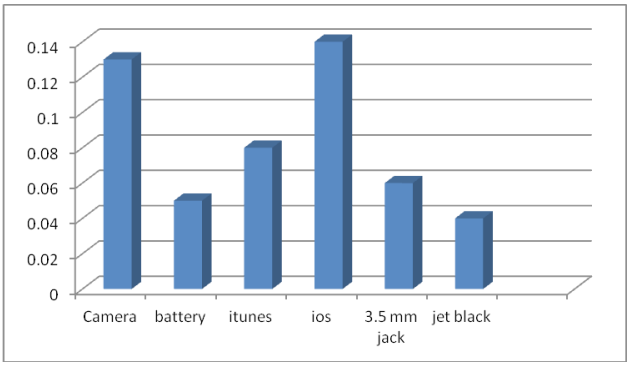

Figure 3. Countrywide concept average.

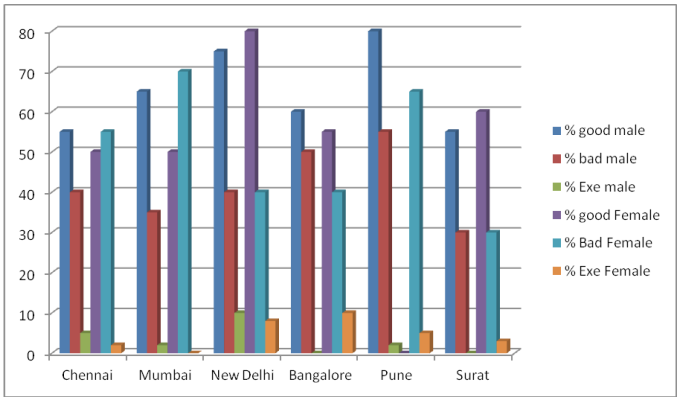

Figure 4. City based gender specific sentiment.

New Delhi shows maximum positive sentiment scores for both genders (female/male), followed closely by Pune and then by Mumbai. This is generally a good sign of diversity in these cities, meaning very diverse user sentiments for the iPhone7 are available. Here % bad, % good and % exe means percentage of marked as "bad", "good" and "excellent" respectively.

5. Conclusion and Future Work

In this research work, a methodology is proposed to identify the sentiment/opinion /popularity of any product/service across female and male users, in several locations. The proposed methodology is a generalized approach and can be used for tweets of any product and from any country with a condition of availability of product related tweets in a good amount. The tool named "NamSor" is used to classify gender of each tweet, SentiWordNet to calculate senti_Score and ConceptNet ontologies to extend concepts are used. Finally in result section, several comparisons are made to validate the accurateness of the proposed approach with the help of graphs. The analysis shows the positive and negative sentiment of male and female users from various locations, towards features of iphone7. It is also analyzed that adding of the concepts from commonsense knowledge increases the performance of proposed system. Now, the future work involves identifying the more relations to extract the concepts and to enhance the quality of concept. Ontologies other than ConceptNet can be explore to improve the concept mining method. To this end, fuzzy based algorithms van be used to enhance the effectiveness of the system in terms of performance.

6. References

1. Erik C, Amir H, Catheine H, Chris E. Common sense computing: From the society of mind to digital intuition and beyond. In: Lecture notes in computer science 5707, Springer. 2009; 252–9.

2. Mullen T, Collier N. Sentiment analysis using support vector machines with diverse information sources. In: EMNLP. 2004; 546–9. PMid:15256037
3. Turney PD. Thumbs u or thumbs down? Semantic orientation applied to unsupervised classification of reviews. ACL-2002. 2002; 417–24.
4. Gelfand B, Wulfekuler M, Punch WF. Automated concept extraction from plain text. In: AAAI workshop on text categorization. 1998; 13–7.
5. Agarwal B, Poria S, Mittal N, Gelbukh A, Hussain A. Concept-level sentiment analysis with dependency-based semantic parsing: a novel approach. Cognitive Computation. 2015; 7(4):487–99. https://doi.org/10.1007/s12559-014-9316-6
6. Hridoy SA, Ekram MT, Islam MS, Ahmed F, Rahman RM. Localized twitter opinion mining using sentiment analysis. Decision Analytics. 2015; 2(1):8. https://doi.org/10.1186/s40165-015-0016-4
7. Nithish R, Sabarish S, Abirami AM, Askarunisa A, Navaneeth Kishen M. An Ontology based Sentiment Analysis for mobile products using tweets. In Fifth International Conference on Advanced Computing. 2013. p. 242–7. https://doi.org/10.1109/icoac.2013.6921974
8. Khanaferov D, Luc C, Wang T. Social Network Data Mining Using Natural Language Processing and Density based Clustering. In IEEE International Conference on Semantic Computing (ICSC). 2014. p. 250–151. https://doi.org/10.1109/icsc.2014.48
9. Viklund A. Free Map Tools. Available from: http://www.freemap-tools.com/. Retrieved December 23, 2016.
10. SNLP Manua. Stanford Typed Dependencies Manual. Available from: http://nlp.stanford.edu/software/dependencies_manual.pdf.
11. SNLP Group. The Stanford NLP Group. Available from: http://nlp.stanford.edu/. Retrieved December 4, 2016.
12. Havasi C, Speer R, Alonso JB. Conceptnet 3; a flexible, multilingual semantic network for common sense knowledge, In: Recent advances in natural language processing. 2007; 27–9.
13. Wang Q-F, Cambria E, Liu C-L, Hussain A. Common Sense knowledge for handwritten Chinese text recognition. Cogn Compu. 2013; 5(2):234–42. https://doi.org/10.1007/s12559-012-9183-y
14. SentiWordNet. Available from: http://sentiwordnet.isti.cnr.it/. Retrieved January 30, 2017.
15. Namsor. Available from: https://github.com/namsor/namsor-api. Retrieved January 30, 2017

Citation:

Amita Jain and Minni Jain

"**Location based Twitter Opinion Mining using Common-Sense Information**",

Global Journal of Enterprise Information System. Volume-9, Issue-2, April-June, 2017. (http://informaticsjournals.com/index.php/gjeis)

Conflict of Interest:

Author of a Paper had no conflict neither financially nor academically.