Detection and Scoring of Internet Slangs for Sentiment Analysis Using SentiWordNet

ARTICLE *in* LIFE SCIENCE JOURNAL · MAY 2014

Impact Factor: 0.17

1 AUTHOR:

Dr. Muhammad Zubair Asghar

Gomal University

38 PUBLICATIONS **34** CITATIONS

SEE PROFILE

Detection and Scoring of Internet Slangs for Sentiment Analysis Using SentiWordNet

Fazal Masud Kundi[1], Shakeel Ahmad[1], Aurangzeb Khan[2], Muhammad Zubair Asghar[1]

[1]Institute of Computing and Information Technology, Gomal University, D.I. Khan, Pakistan
[2]Institute of Engineering and Computer Sciences, University of Science and Technology Bannu, Pakistan
fmkundi@gu.edu.pk, zubair_icit@yahoo.com

Abstract: The online information explosion has created great challenges and opportunities for both information producers and consumers. Understanding customer's feelings, perceptions and satisfaction is a key performance indicator for running successful business. Sentiment analysis is the digital recognition of public opinions, feelings, emotions and attitudes. People express their views about products, events or services using social networking services. These reviewers excessively use Slangs and acronyms to express their views. Therefore, Slang's analysis is essential for sentiment recognition. This paper presents a framework for detection and scoring of Internet Slangs (DSIS) using SentiWordNet in conjunction with other lexical resources. The comparative results show that proposed system outperforms the existing systems.

[Fazal Masud Kundi, Shakeel Ahmad, Aurangzeb Khan, Muhammad Zubair Asghar. **Detection and Scoring of Internet Slangs for Sentiment Analysis Using SentiWordNet.** *Life Sci J* 2014;11(9):66-72]. (ISSN:1097-8135). http://www.lifesciencesite.com. 10

Keywords: Sentiment Analysis; Slangs; Microblogs; Social Media; Semantic Orientation

1. Introduction

The Web 2.0 has dramatically changed people's communication style. It is a great move toward more community oriented, highly collaborative, interactive and responsive Web (Socialmedialeap, 2014). Today we are not only using the Internet but we are part of this global network. Revolution of the social networking is the direct significant impact of Web 2.0. Social media sites became a world largest virtual community, where people express their views about products, events and services, anytime from anywhere (Yang WC et al., 2013). These views have great impact on community thinking and decisions. Information and communication technology (ICT) have made radical changes to various fields such as business, commerce, economy and banking (Ghods M et al., 2014).The statistics (Loyalrewards, 2014) show that 70% customers trust online reviews, like a personal recommendations, and 80% of consumers have changed their decisions about buying a product on the basis of negative reviews found on Web.

Mountains of online information in shape of reviews and blogs are generated every minute with unprecedented speed and size. These reviews contain opinions, comments, and feelings of users, which is the rich source for sentiment analysis. Sentiment analysis or opinion mining, is the computational study of people's sentiments, attitudes, appraisals, and emotion toward physical or logical entities, such as products, events or services (Liu B, 2012). Sentiment analysis received considerable attention of researchers over past few years (Xu T and Peng Q, 2012) due to the popularity of social networking services such as Facebook, Twitter, MySpace and LinkendIn.

Slangs abbreviations are mostly used by the Internet users in their messages as shown in table 1. Slang is a type of language of non-standard words and phrases (Wikipedia, 2014), such as GR8, SMH, CHALE and XOXO. The primary motivation behind the using of Slang words is its usefulness, because usually easy for other to interpret and save a lot of time. Large number of Slangs with positive or negative sentiments are used in chat, Twitter and Facebook messages (Asghar MZ et al., 2014). It has become very important to detect, translate and identify Slang's polarity for determining the sematic orientation (SO) of the entire review.

This paper presents a framework to detect and score the Slang words for sentiment analysis. The remaining paper is structured as follow. Related work is described in Section 2 followed by the proposed framework in section 3. Section 4 presents experimental setup, section 5 evaluates the results and conclusion is given in section 6.

Table 1. Tweets Containing Slangs

S/No.	Tweet
1	My new Samsung Galaxy Tab 3: D @chrisfosterelli, you are the best! **Xoxo**
2	first day of my shoot..butterfliez in my stomach wish me luck guys..n have a **gr8 gr8** day love..
3	Ur a dream crusher u crushed my dream again **smh** demon:
4	Dear iPhone I have typed **hahaha** like a million times yet you continually give me **hagaha bahaha** gagaha

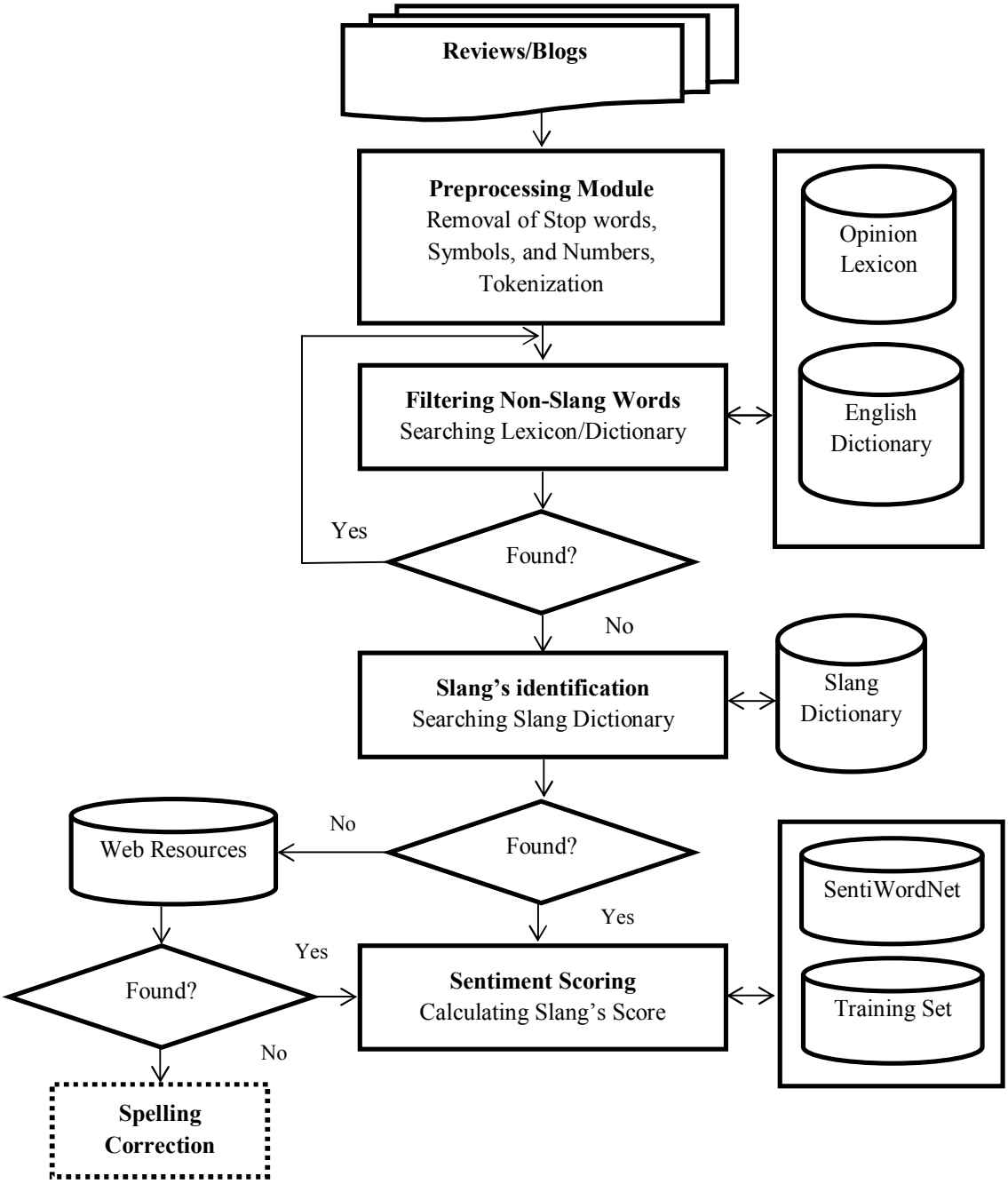

Figure 1. Slang's Detection and Scoring Framework

2. Related Work

Sentiment analysis tasks can be classified on the basis of their using levels i.e. document level (Turney PD, 2002;Pang B and Lee L, 2008; Othman M et al., 2014), sentence level (Hu M and Liu B, 2004; Kim SM and Hovy E, 2004) and word or phrase level (Kumar A and Sebastian TM, 2012). In document level SO can be accumulated from the words or phrases to find out overall SO of a particular sentence or review. In sentence level the target review/blog is splitted into sentences and SO of each sentence is calculated using lexical or statistical methods (Khan A et al., 2011). At word level, mostly adjectives or adverbs are used that have SO (Andreevskaia A and Bergler S, 2006; Esuli A and Sebastiani F, 2006). The given word is classified as a positive, negative or neutral based on their semantic orientation. Feature-based model (Liu B et al., 2005) extracts the sentiment words from text and classifies them accordingly.

Text sentiment classification approaches can be classified into two i.e. machine learning and score-based approaches (Wang S et al., 2011; Chen LS et al., 2011). Machine learning techniques require training documents (annotated) for classifier training. The score-based methods treat several attributes of an entity as a sub-problem. In the first phase these attributes are classified into positive and negative classes using semantic properties (Turney PD and Littman ML, 2003). In second phase the sentiment score for the entire document is calculated by summing scores of all attributes. If total sum is positive then positive sentiment is assigned to the document otherwise negative sentiment is assigned (Yessenalina A et al., 2010; Turney PD and Littman ML, 2002). Sentiment scores of Slangs for the given text is presented in (Manuel K et al., 2010). Their approach is based on (Martineau J et al., 2009; Tokunaga T and Makoto I, 1994) work. Some researchers (Nielsen FA, 2011) have developed an opinion lexicons with sentiment score for each word including Slang words. But it covers a small portion of Slang words. Slang words are also frequently used in community-based question answering (cQA) and a rich source for sentiment analysis (Amiri H and Chua TS, 2012). Limited work is carried out on analysis of Internet Slangs for sentiment analysis.

This research work follows score-based approach for scoring Internet Slangs using SentiWordNet (SWN) (Esuli A and Sebastiani F, 2006) in conjunction with other lexicons and dictionary resources.

3. Proposed Framework

The proposed framework for detection and scoring of Slangs is depicted in figure 1. This framework consists of four major modules.

3.1 Preprocessing Module

All preprocessing tasks related to reviews/blogs are performed in this module. Detail of these tasks is given in section 4.3.

3.2 Filtering Module

This module filters all non-slang words before starting the detection of Slang words to save the "running time" of the detection process. The word is considered as a "non-Slang" if found in an opinion lexicons or dictionary. If the word is not "non-Slang" word then it is passed to the Slang identification module for further processing.

3.3 Slang's Identification

In this module slang dictionary is searched for finding Slang and its definition. If found it is passed to the sentiment scoring module otherwise Web is searched for its definition (meaning). Currently our Web search for slangs is confined to Web-based Dictionary (Urban, 2014). If the Web returns positive response the Slang is scored. In case of negative result it is treated as a misspelled word. This framework corrects the basic spelling errors such as word with repeated letters.

3.4 Sentiment Scoring

This module calculates the sentiment score for the given Slang word. In most of the cases when a Slang definition is retrieved from Web, it contains noisy text and requires preprocessing as shown in table 4. The semantic score of the pre-processed text (Slang's definition) is retrieved automatically from SWN. SWN associates three numeric values with each synset of the Wordnet (Miller GA et al., 1990) i.e. *pos(w), neg(w)* and *obj(w).* Sum of all three values is equal to 1. Each entry in SWN has multiple senses. Average of *pos(w), neg(w)* and *obj(w)* scores for each sense is calculated according to part of speech (POS). To make the score more contextual, a fraction of positive, negative or neutral reviews which contain the target Slang is added to the SWN score. If the *obj(w)* is less than threshold value (0.5) the word is considered as positive or negative. Positive difference between *pos(w)* and *neg(w)* means positive word otherwise it is negative. The overall scoring formula and its components are shown below.

$$pscore(w) = \sum_{i=1}^{n} pos(s_i)/n \quad (1)$$

$$nscore(w) = \sum_{i=1}^{n} neg(s_i)/n \quad (2)$$

$$oscore(w) = \sum_{i=1}^{n} obj(s_i)/n \quad (3)$$

Where *pscore*(*w*), *nscore*(*w*) and *oscore*(*w*) represent positive, negative and objective polarity score of all synsets for word *w, s_i* is the score of *ith* synset and *n* is the total number of synsets. If sum of *pscore* and *nscore* is greater than threshold value (0.5) for all words in the Slang's definition then formula in (4) is used to calculate the overall sentiment score for the given slang.

$$Score(slang) = (\sum_{i=1}^{m} pscore(w_i) + pf) - (\sum_{i=1}^{m} nscore(w_i) + nf) \quad (4)$$

Where w_i represents *ith* word and *m* is the number of words in the Slang definition. The term *pf* and *nf* denote the fraction of positive and negative reviews respectively that contain the target slang. Scoring algorithm is presented in figure 2.

4. Experimental Setup

This section describes the experimental setup for this research work in detail.

4.1 Lexical Resources Used

Following lexical resources were used in this research work: (i) General purpose opinion lexicon (Hu M and Liu B, 2004), which contains 1967 positive and 4783 negative sentiment words. (ii) Dadvar opinion lexicon (Dadvar M et al., 2011), which classifies English words into 136 positive, and 109 negative sentiment words. (iii) Wordnet (Miller GA et al. 1990), which is a lexical repository for English language. It is comprised of 155,287 words and 117,659 synsets, also called synonyms. (iv) SentiWordNet, another lexical resource and an extension of Wordnet. It associates each Wordnet synset with three numerical scores i.e. positive, negative, and objective. These numerical values exist between 0 and 1 inclusive, with sum equal to unity. (v) English dictionary (Wordlist, 2013), which contains 79768 words was used for filtering non-slang words and basic spelling correction. (vi) More than 5000 Slangs (acronyms) were collected from the Web (Slang Dictionary, 2014) and compiled with their translation for scoring. (vii) Urban dictionary, which is web based dictionary founded in 1999, and contains more than seven million definitions. It is one of the best Web site among social media users. Our Slang's detection and translation process ends at Urban dictionary.

4.2 Dataset

We performed experiments on two datasets of microblogs: (i) Dataset1, which contains 13186 un-annotated Tweets. This dataset was used to test the framework in detection of Slang words without scoring it. (ii) Dataset2, which contains 150 manually annotated (positive, negative or neutral) Tweets. The purpose of this dataset is testing framework in detection, scoring of Slang words and its impact on sentiment classification performance. Both Datasets were collected by using Twitter streaming API. Datasets statistics are shown in table 2.

Input: Online Reviews/Blogs
Output: Slang and its polarity
Function_Slang(text**)**
1. ptext = preprocessor(text)
2. tokens = tokenize(ptext)
3. **For** token in tokens
4. Filtering (skipping non-Slang words)
5. Search Slang dictionary
   If found **Then**
   Get Slang's meaning
   Process and determine its polarity
   Goto step 7
6. Search Web for Slang's detection
   If found **Then**
   Get Slang's meaning
   Process and determine its polarity
   Else
   Correct Spelling/Discard
7. **Next**

End Function

Figure 2. Slang's Detection and Scoring Algorithm

Table 2. Statistics of Datasets

Dataset	Tweets	Tokens	Unique Tokens
1 (Un-annotated)	13186	126998	29126
2 (Annotated)	150	890	628

4.3 Preprocessing

To avoid incorrect and misleading results, data must be preprocessed before using it. All basic preprocessing tasks (Asghar MZ et al., 2014) were performed such as removal of irrelevant Tweets, stop words, numbers, URLs and hash tags. In addition to the above following preprocessing tasks were also performed:

- Word tokenization.
- Filtering non-slang words using lexicons and dictionary resources.
- Tokens having length less than 2 and greater than 10 are ignored.
- Preprocessing of Slang's text (meaning) before scoring.
- Unique tokens (29126) were further filtered leaving 7046 as a candidate for Slang words.

Table 3. Polarity of Slang Words

S/No.	Slang	Meaning	Score	Orientation
1	Alr	Alright	0.25	Positive
2	Chale	Disagreement or Disapproval	-0.0928	Negative
3	Coolio	Cool	0.080338	Positive
4	Damn	Condemn/Disbelief	-0.16477	Negative
5	Gonna	Want to go	0.023256	Neutral
6	gr8	Great	0.30814	Positive
7	Haha	Laughing	0.011628	Neutral
8	Hamm	Powerful	0.198863	Positive
9	Happs	Happy	0.5625	Positive
10	Hehehe	Laughing	-0.1875	Negative
11	Hurr	Greeting	0.523256	Positive
12	Ll	Laughing Loudly	0	Neutral
13	Notta	Not	-0.67262	Negative
14	Rofl	Rolling On Floor Laughing	0.008854	Neutral
15	Smh	Shaking My Head	-0.0671	Negative
16	Tbh	To Be Honest	0.261628	Positive
17	Wanna	Want to	0.011628	Neutral
18	Xoxo	Hugs and kisses	0.137839	Positive

4.4 Performance Evaluation

Confusion matrix (CF) (Provost FJ et al., 1998) also called error matrix is used to present the system performance. CF presents the actual and predicted results in form of TP, TN, FP and FN, which stands for true positive, true negative, false positive and false negative respectively. Researchers use various performance measures for system evaluation such as: (i) Precision (Olson DL and Delen D, 2008) also called positive predicted value, measures the correctness of the model. Higher precision indicates less FP. Mathematically it is defined as:

$$Precision, p = \frac{TP}{TP+FP} \quad (5)$$

(ii) Recall (Olson DL and Delen D, 2008) also known as sensitivity, measures positive cases classified correctly by the model. Large recall value means few positive cases miss classified as a negative. Recall can be calculated using the following formula.

$$Recall, r = \frac{TP}{TP+FN} \quad (6)$$

(iii) F-Score or F-measure (Olson D L and Delen D, 2008) is based on both precision and recall. It is calculated as follow:

$$F - Score = \frac{2rp}{r+p} = \frac{2TP}{2TP+Fp+FN} \quad (7)$$

5. Results and Discussion

We performed experiments on microblogs data which is the rich source for analysis of Slangs, acronyms and emoticons. The unique tokens (29126) of dataset 1 were filtered leaving 7046 tokens as a candidate for Slang words. 63% tokens were successfully identified as a Slang words. The 167 tokens in remaining of 37% were corrected by the basic spell checker and filtered. The spell checker checks the spelling at the last if the system cannot recognize the token as a Slang. If the spell checker fails then the token is considered as garbage and discarded.

This framework can be used to generate the opinion lexicon for Slang words. It almost detects any Slang in reviews or blogs because this framework is extendable. English words list called AFINN-111 (Nielsen FA, 2011) contains 2477 opinion words including Slangs. This list-based approach performed better than many other comprehensive lists in calculating sentiment of tweets (Nagy A et al., 2012). However the list covers a small number of Slangs. Slangs detected by our system shown in table 3, were searched in the above

list, only three Slangs (Damn, Haha, Rofl) were found there.

Table 3 shows the score of some common Slang words along with their SO. A large number of Slangs/acronyms were detected in the target dataset. Most of the Slangs were ignored due to its irrelevancy or having no positive or negative sentiments. For example BBM (Black Berry Messenger), DWI (Driving While Intoxicated), IBM and LEXUS (Japanese Car) are irrelevant words. As Urban Dictionary defines millions of terms in a multiple ways, so it is very difficult to filter irrelevant words and definitions. In most cases we have taken the first definition (meaning) of the Slang. Table 4 shows multiple definition of the slang word "CHALE". In majority of cases multiple definitions conflict with each and other. It is very important to determine Internet Slang's polarity before sentiment classification of reviews or blogs. Table 5 shows impact of Slang analysis on sentiment classification. Comparative performance for sentiment classification is shown in table 6.

Table 4. Slang Definitions (Urban Dictionary)

Slang: xoxo (First definition out of 14)
Hugs and kisses O=Hug X=Kiss If you look at each letter like it was representing two people from a bird's eye view, the "O" represents the arms of those persons hugging each other while the "X" is evocative of two people kissing each other.
Slang: chale (First 5 definitions out of 7)
Def#:1 = THIS IS THE REAL DEFINITION. A term to show disagreement or disapproval of something or some idea. Means the same thing as "hell no" or "hell na."
Def#:2 = Chicano term used for disagreement
Def#:3 = A spanish word that is used by latinos when they are saying "forget about it" or "no way".
Def#:4 = Chale (GHANA) *Alternative Spellings: Charlie, Charle *pronounced: "Char-lay/Chaa-lay"
Def#:5 = Spanish slang term usually used by Mexicans/Chicanos for " damn ! "; " oh my god "; or " yeah right " it all depends sometimes it can mean shut up.

Table 5. Impact of Slangs on Sentiment Classification Performance

	Result (%) A = Accuracy, P = Precision, R = Recall F = F-Score, S1/S0 = With/Without Slangs							
	Overall		**Positive**		**Negative**		**Neutral**	
	S1	**S0**	**S1**	**S0**	**S1**	**S0**	**S1**	**S0**
A	87	85	89	85	95	93	65	65
P			98	97	74	72	81	68
R			89	85	95	93	65	65
F			93	91	83	81	72	67

Table 6. Comparative Performance

	Method	**Precision**	**Recall**	**F-Score**
Positive	(Amiri et al. 2012)	72.34	**96.59**	82.72
	DSIS	**97.50**	88.64	**92.86**
Negative	(Amiri et al. 2012)	**96.45**	58.80	73.06
	DSIS	74.07	**95.23**	**83.33**

6. Conclusion and Future Work

In this paper, we presented a framework for detection and scoring of Slang words for sentiment analysis. We achieved satisfactory results by using this framework on microblogs datasets. This framework can be extended by enhancing the preprocessing, detection and scoring modules. Comprehensive context-sensitive and sentiment-aware spell corrector is also proposed for future work.

References

1. Amiri H. and Chua TS. Mining slang and urban opinion words and phrases from cQA services: an optimization approach. In Proceedings of the fifth ACM international conference on Web search and data mining 2012;193-202.
2. Andreevskaia A and Bergler S. Mining WordNet for a Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses. In EACL 2006;6:209-215.
3. Asghar M Z, Khan A, Ahmad S and Kundi FM. A Review of Feature Extraction in Sentiment Analysis. Journal of Basic and Applied Scientific Research 2014;4(3):181-186.
4. Chen LS, Liu C H and Chiu HJ. A neural network based approach for sentiment classification in the blogosphere. Journal of Informetrics 2011;5(2):313-322.
5. Dadvar M, Hauff C and de Jong FMG. Scope of negation detection in sentiment analysis. In Dutch-Belgian Information Retrieval Workshop 2011;16-19.
6. English Word Lists for Spell Checkers, http://www.curlewcommunications.co.uk/wordlist.html <Accessed:12/2013.
7. Esuli A and Sebastiani F. Sentiwordnet: A publicly available lexical resource for opinion mining. In Proceedings of LREC 2006;6:417-422.
8. Ghods M et al. Evaluation of the Effective Factors on Online Internet Usage in Organizations. Life Science Journal 2014;11(1):58-63
9. Hu M and Liu B. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining 2004;168-177.
10. Importance of having a Web 2.0 Blog, http://www.socialmedialeap.com/blog-design/ <Accessed:02/04/2014.

11. Internet Slang Dictionary and Translator, http://www.noslang.com/ <Accessed:01/2014.
12. Khan A, Baharudin B, and Khan K. Sentiment Classification Using Sentence-level Lexical Based Semantic Orientation of Online Reviews. Trends in Applied Sciences Research 2011;6(10):1141-1157.
13. Kim SM and Hovy E. Determining the sentiment of opinions. In Proceedings of the 20th international conference on Computational Linguistics 2004;1367.
14. Kumar A and Sebastian TM. Sentiment analysis: A perspective on its past, present and future. International Journal of Intelligent Systems and Applications (IJISA) 2012; 4(10): 1-14.
15. Liu B, Hu M and Cheng J. Opinion observer: analyzing and comparing opinions on the web. In Proceedings of the 14th international conference on World Wide Web 2005;342-351.
16. Liu B. Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies 2012;5(1):1-167.
17. Manuel K, Indukuri KV and Krishna PR. Analyzing Internet Slang for Sentiment Mining. In Second Vaagdevi International Conference on Information Technology for Real World Problems 2010;9-11.
18. Martineau J and Finin T. Delta TFIDF: An Improved Feature Space for Sentiment Analysis. In Proceedings of the Third AAAI Internatonal Conference on Weblogs and Social Media 2009;258-261.
19. Miller GA, Beckwith R, Fellbaum C, Gross D and Miller KJ. Introduction to Wordnet: An on-line lexical database. International journal of lexicography 1990;3(4):235-244.
20. Nagy A, Valley C M S and Stamberger J. Crowd sentiment detection during disasters and crises. In Proceedings of the 9th International ISCRAM Conference 2012;1-9.
21. Nielsen FA. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs, arXiv preprint arXiv 2011;103(2903).
22. Olson DL and Delen D. Performance Evaluation for Predictive Modeling. In Advanced data mining techniques. Springer-Verlag Berlin Heidelberg 2008;137-139.
23. Othman M, Hassan H, Moawad R, and El-Korany A. Opinion Mining and Sentimental Analysis Approaches: A Survey. Life Science Journal 2014;11(4):321-326.
24. Pang B and Lee L. Opinion mining and sentiment analysis. Foundations and trends in information retrieval 2008;2(1-2):1-135.
25. Provost FJ, Fawcett T and Kohavi R. The case against accuracy estimation for comparing induction algorithms. In ICML 1998; 98:445-453.
26. Tokunaga T and Makoto I. Text categorization based on weighted inverse document frequency. Department of Computer Science, Tokyo Institute of Technology, Technical Report 94-TR00001 1994.
27. Turney P D and Littman M L. Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems (TOIS) 2003;21(4):315-346.
28. Turney PD and Littman ML. Unsupervised Learning of Semantic Orientation from A Hundred-billion-word Corpus. Technical Report EGB-1094. National Research Council Canada 2002.
29. Turney PD. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th annual meeting on association for computational linguistics 2002;417-424.
30. Urban Dictionary, http://www.urbandictionary.com/ <Accessed:01/2014
31. Wang S, Li D, Song X, Wei Y and Li H. A feature selection method based on improved fisher's discriminant ratio for text sentiment classification. Expert Systems with Applications 2011;38(7):8696-8702.
32. Why you should care about online reviews? http://loyalrewards.com/images/Online-Review-Infograph-MT.pdf/ <Accessed:20/03/2014.
33. Wikipedia, http://en.wikipedia.org/wiki/Slang/ <Accessed:01/2014.
34. Xu T, Peng Q and Cheng Y. Identifying the semantic orientation of terms using S-HAL for sentiment analysis. Knowledge-Based Systems 2012; 35:279-289.
35. Yang WC, Chen KC, Wu C L and Chang CM. A Behavioral Model of Consumer's Usage of Travel Blogs. Life Science Journal 2013;10(2):1468-1474.
36. Yessenalina A, Yue Y and Cardie C. Multi-level structured models for document-level sentiment classification. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing 2010;1046-1056.
37. Asghar, M. Z., RahmanUllah, B. Ahmad, A. Khan, S. Ahmad and I. U. Nawaz, 2014. Political miner: opinion extraction from user generated political reviews. Sci.Int(Lahore), 26(1): 385-389.