

Sentence Level Sentiment Analysis Using Urdu Nouns

Faiza Hashim

Department of Computer Science,
University of Peshawar, Pakistan
fl.shah@yahoo.com

Mohammad Abid Khan

Department of Computer Science
University of Peshawar, Pakistan
mabid6@gmail.com

Abstract

Some pioneer effort has been made for Urdu sentiment analysis and a lot more is needed to explore this field. The goal of this research study is to contribute in the field of sentiment analysis for Urdu language by extending the scope of sentiment expressions from adjectives to nouns in the domain of news. In this study, a sentiment analyzer for Urdu opinionated data is developed which works at sentence level to analyze public opinions given on news headlines. A lexicon based approach is used by the developed sentiment analyzer which exploits Urdu lexicon of adjective and noun class expressions. Urdu opinionated corpus construction and Urdu sentiment lexicon construction are part of this study. The main novelty of this study is the use of nouns as sentiment carriers along with adjectives. Experimental results are used to evaluate and show better performance of the system compared to previous approaches used for Urdu sentiment analysis.

1. Introduction

Sentiment analysis (SA) is the fastest growing field of Natural Language Processing and Text Mining under the umbrella of Artificial Intelligence (AI). The field of sentiment analysis emerged from human behavior of decision making by consultation and asking friends or family about their opinions in daily life. Examples are buying a product or a service, observing public response about a specific law or policy approved by government; increasing demands of customers and employees performance by their companies and political reviews during elections.

The increasing availability and use of online resources for opinion sharing on social media, such as news, blogs, review sites, has greatly facilitated the decision making process of several parties such as

customers, governments and companies. A huge volume of opinionated data is entered daily by exploiting the World Wide Web over the internet in different local and regional languages along with English. As compared to English language, these local and regional languages are observed as resource poor languages as data and tools are scarcely available. Developing automated sentiment analyzers for such resource poor languages is a challenging task these days.

Sentiment analysis can be performed on three different levels, depending on the nature of the data and choice of the users. These three levels of analyses are:

- Document level sentiment analysis
- Sentence level sentiment analysis
- Aspect level sentiment analysis

The document level sentiment analysis classifies the overall sentiments of a document as being positive or negative. A single result, positive or negative, holds for the whole document which fails to detect sentiments about individual aspects of the topic [1].

The sentence level sentiment analysis classifies each sentence as positive, negative or neutral. At this level, sentence must be a separate unit, expressing a single opinion.

In aspect level, the data for analysis comprises those entities which have one or more attribute(s)/feature(s). The opinions on these features can be expressed by different opinion holders on a single feature. The aspect level sentiment analysis produces a feature-based opinion summary of multiple reviews, classifying sentiments of each attribute/feature as positive or negative.

The two main approaches used for the problem of sentiment analysis are machine learning approach and lexicon-based approach. In machine learning approach, the text classification is performed by training the classifier on labeled data. Any text classification algorithm can be used for this approach, such as Naïve

Bayes, SVM (Support Vector Machine), Maximum Entropy, Multilayer Perception, and Clustering.

The lexicon-based approach works on sentiment words dictionary, i.e. the lexicon. The lexicon consists of predefined list of sentiment words associated with their polarity and intensities. The lexicon varies according to the context and it is difficult to maintain a unique lexicon which can work in different contexts. As compared to machine learning approach, lexicon-based approach is considered as a simple approach as it does not require labeled data [2].

Corpus-based and Dictionary-based approaches are also used by some researchers. Corpus-based approaches find co-occurrence patterns of words to determine the sentiments of words or phrases. Dictionary-based approaches use synonyms and antonyms in WordNet to determine word sentiments based on a set of seed opinion words [3].

The role of adjectives as sentiment carrier is central and cannot be denied. However, in some specific situations e.g. news domain, subjective nouns can also work as sentiment carriers and can increase the performance of the system as nouns appear frequently in news data. Example (1), (2), (3), (4) and (5), taken from [4], show the importance of nouns as sentiment expressions. In these examples, the underlined terms are nouns, which are used to classify the opinions as positive or negative.

- (1) بلاول کی نقل و حرکت کو محدود کرنے کی سازش ہے۔
Bilawāl ki naql o harkath ko mehdud karne ki sāzish hai.
“This is a conspiracy to restrict Bilawal’s mass campaign”.
- (2) کچھ نہیں سب ڈرامے بازی ہے۔
Kuch nahi sab dram e bāzi hai
“It is nothing but a stage act”.
- (3) اللہ حامد میر کو جلد صحتیاب کرے۔
Allah Hāmid mir ko jald sehathyāb kary
“May Allah make Hamid Mir well soon”.
- (4) میں جماعت اسلامی کے سیاسی کردار سے مکمل اتفاق کرتا ہوں
Mein Jamāt-i Islami ke siyāsi kirdār se mukam’al ithifāq kartha hun
“I fully endorse Jamat e Islami’s political action”.
- (5) ایسے لوگ صدیوں میں پیدا ہوتے ہیں صائم شہادت کے مرتبے پر فائز ہونی ہیں صائم کو سلام۔
Aisy log sadiyo mein paida hothy hain. Sāimā shahādad ke marthaby par fāiz hui hain. Sāima ko salām

“Such people are born in centuries. Saima embraced martyrdom. We pay our tribute to her”.

2. Motivation

While dealing with sentiment analysis for English language and other international languages, this field appears as an already greatly explored one. The reason is that a lot of tools and resources are available and much research is done and is going on. However, moving to regional languages, sentiment analysis appears as a newly emerging field.

Urdu is the national language of Pakistan and is spoken by 60.5 million speakers in the Indian subcontinent [5]. Urdu is written with Arabic script from right to left and the recommended writing style is Nastalique [6]. The research progress in such resource poor languages is a challenging task. The main motivation for the present research work is to facilitate sentiment analysis for Urdu language and develop automated sentiment analyzer to mine Urdu opinionated data using nouns as sentiment expressions along with adjectives.

3. Related Work

Sentiment analysis is a text classification technique which is used to classify opinions as positive, negative or neutral. Usually, this classification starts at word or phrase level [7, 8, and 9] and moves to sentence level [10, 11, 12, 13, 16 and 18] and to document level [14, 15].

Some effort was also done to draw a comparison for better sentiment analysis between different levels of analyses [16, 17].

In some cases the output of one level is used as input to other levels, as reported in [17, 18, 19].

The lexicon based approach works on sentiment words dictionary, i.e. the lexicon [14, 35].

In sentiment lexicons are available which include WordNet [18, 20] and SentiWordNet [21, 22, and 23]. In the field of sentiment analysis, adjectives are traditionally considered the center of attention as adjectives are sentiment carriers which are used to determine the polarity of given text [9, 15, 26, 29, 35, 36 and 37]. Some effort has been made to consider other parts of speech along with adjectives as sentiment expressions. The work includes the use of adjectives and verb class information [18, 38], adjectives and adverbs [15, 39], “adjective verb adverb” framework [33] and non-effective adjectives and adverbs [34].

While dealing with Urdu sentiment analysis, only adjectives are considered as the sentiment expressions in opinionated text [9, 25, 26, 28 and 29]. The tools and automated systems developed so far for English language cannot be used exactly for Urdu data due to the vast orthographic, morphological and grammatical differences between both the languages [9].

Mukund and Srihari made a pioneer effort in Urdu sentiment analysis and developed a classifier to distinguish subjective sentences from objective sentences of Urdu language [24].

Urdu sentiment lexicon was developed by [9] for the first time to performed lexicon based sentiment parsing.

The work reported in [25] directed the core issue in analyzing sentiment i.e. negation handling and handled negation at phrase level.

Sayed et al. highlighted the importance of adjectival phrases in sentiment analysis and used the term “SentiUnits” for expressions containing sentiment information [26].

Mukund and Srihari used the method of structural correspondence learning (SCL) to transfer sentiment analysis learning from Urdu newswire data to Urdu blog data exploiting the code switching and code mixing techniques [27].

The work reported in [28] performed lexicon based sentiment analysis at aspect level and associated targets with sentiment expressions and the result was a better performance compared to their earlier work [9].

A bilingual lexicon was developed by [29] using bilingual datasets (English and Roman Urdu) and performed lexicon based approach for bilingual sentiment analysis of tweets.

4. Methodology

The current study uses a lexicon based approach for Urdu sentiment analysis which works at sentence level to categorize public opinions posted on news headlines [4]. Although machine learning approaches give promising results but due to the unavailability of Urdu opinionated labeled data, the current research adopted a lexicon based approach. Sentence level sentiment analysis gives better performance than document level sentiment analysis because at sentence level sentiment analysis each opinion is analyzed and considered for being positive or negative, however in document level sentiment analysis a single result holds for whole document. Considering the nature of data, aspect level sentiment analysis cannot be used as in news data no

direct mapping of entities with their attributes exists. Therefore sentence level sentiment analysis is more useful for this study. Urdu lexicon and Urdu corpus construction are additional tasks of this study. Figure 1 shows a complete step by step workflow for Urdu sentiment analyzer.

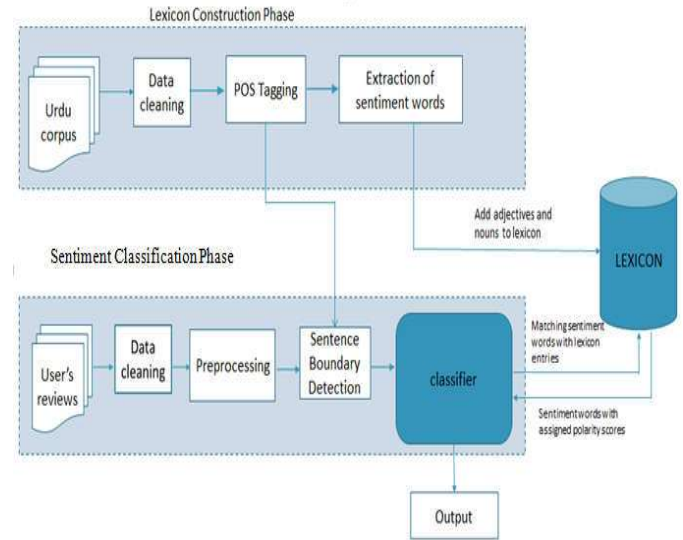


Figure 1: Urdu Sentiment Analyzer Flow Chart

The proposed Urdu Sentiment Analyzer works in two phases.

- Lexicon Construction Phase
- Sentiment Classification Phase

The lexicon construction phase involves preprocessing the corpus data and extraction of sentiment words from corpus and construction of lexicon using these words.

The sentiment classification phase performs the actual function of sentiment analysis i.e. mining the opinions as positive, negative or neutral. The input to this phase is taken from previous phase and consults the lexicon for making decision.

4.1. Sentence Boundary Identification

The proposed method is performed at sentence level. At this level, sentence must be a separate unit expressing a single opinion given by a single opinion holder. Therefore, a “#” sign is manually inserted at the end of each opinion in the collected corpus, to separate it from other opinions and make it a complete single unit.

4.2. Urdu Corpus Creation and Data Cleaning

A corpus of public opinions on Urdu news headlines has been collected to be used for sentiment analysis [4]. About 1000 opinions (positive, negative and neutral) have been collected from different domains including current affairs, politics, sports, health and entertainment from daily jang blog [4]. The corpus collected so far is in raw form; therefore, the corpus is preprocessed by removing unwanted information (opinion holder's names, addresses, date of posting and dashed lines), soft and hard spaces and opinions/words given in foreign language.

4.3. Parts of Speech Tagging

Part of Speech (POS) tagging is the process of assigning a tag showing the part of speech of a given word e.g. the symbol NN for a noun, ADJ for an adjective and PN for proper nouns. Part of speech tagging plays a vital role in enabling the algorithm to extract and match information using POS tags associated with the words.

After preprocessing the corpus, the data is tagged using tagger for Urdu data developed by [30]. Until this implementation work, the above mentioned tagger [30] was not officially released and was not publically available for personal use. Therefore, the corpus tagging has been performed by these authors of [30] themselves as a response to a request. This tagger shows the performance up to 88.74% accuracy level. However, proper noun (PN) tags are corrected manually to achieve better accuracy, which included the pronoun.

4.4. Lexicon Construction

Lexicon construction is an important phase of the proposed algorithm. The lexicon is constructed from the collected opinionated Urdu corpus. The lexicon consists of two tables i.e. one for nouns and second for adjectives. Words having NN (noun) or ADJ (adjective) POS tags are extracted from corpus and are transferred to their appropriate tables in the lexicon.

Frequencies are assigned to words in the lexicon. Words of lexicon are arranged according to the descending order of their frequencies. The importance of this arrangement of frequencies is to speed up the execution of algorithm as higher frequency words will appear in the start of the table and hence will be matched first. Polarities are also assigned to lexicon entries. A two scale polarity is used i.e. +1 for positive entry and -1 for negative entry. However the absence of adjective and nouns in a sentence gives a neutral

opinion. Due to scarcely availability of Urdu opinionated data, the lexicon constructed in this work contains a total number of 316 distinct nouns and 133 distinct adjectives.

After the completion of the lexicon construction phase, the algorithm starts with the sentiment classification phase which performs the actual functionality of the sentiment analyzer. The detail of each step of this phase is given in the following sections.

4.5. The Classifier

A classifier is a programming module of sentiment analyzer which classifies the given opinion as being positive, negative or neutral. At this stage, the classifier takes an opinion as input, which is preprocessed with sentence boundaries identified; and examine this opinion for adjectives and nouns. If any text is identified with an adjective (ADJ) or noun (NN) tag, the classifier takes this piece of information and consults the appropriate table in the lexicon. If the text is tagged NN (or ADJ) then the item is searched in the nouns (or adjectives) table to find its matching entry in the lexicon. If match is found, its polarity is extracted.

Two counters are used for storing polarity value of words i.e. P variable for positive polarity and N variable for negative one. The polarity counter is incremented when its respective polarity is extracted. However, the polarities assigned in the lexicon can be altered by the presence of negation words, which need to be carefully handled. After solving the negation (Section 4.6.1), the appropriate counters are incremented. The entries are matched with lexicon and polarities are extracted until the “#” symbol is encountered, indicating the end of opinion. Finally, both the counters are compared and result is given depending on the value of the counters, i.e. if P counter value is greater than N counter value then the algorithm gives a “positive opinion” result otherwise “negative opinion”. The classifier gives a “neutral opinion” for opinions which lack any adjective/noun or both.

However, the situation where P and N values are the same (i.e., $P = N$), then the algorithm considers the N variable and give a negative result. The reason behind this decision is the strong influence of negative opinion words over the positive opinions by examining the opinions in the collected corpus. Examples (6) and (7) justify this action [4].

(6) یہاں اچھے کام پر سزا ملتی ہے۔

Yaha achy kām par sazā milthi hai

“Good work is given a negative here”.

(7) قادری صاحب ناکامی مبارک، عوام کو بے وقوف بنانے کا شکر ہے۔

Qādri sahib nākāmi mubārak, awām ko bevaqvaf banāny ka shukriyā

“Congratulations on failing Qadri Sahib. Thank you for fooling the public”.

In above mentioned examples the underlined terms are sentiment words and the presence of equal number of positive and negative sentiment words ($P = N$) give negative results.

4.5.1. Handling Negation. Negation words alter the polarity of associated sentiment words. In Urdu language, ‘نہیں’ and ‘نہ’ act as negation words. The position of the negation words is very important. The current classifier is considering the negation words at one word distance i.e. before or after the target item. If negation word is present in any of these locations then the polarity value is altered otherwise it is taken as original one.

The negation at one word difference is chosen for correct sentiment classification. Exceeding the one word distance in most opinions gave wrong results. Consider the following examples [4] for this justification.

(8) نہیں نواز حکومت کی یہ پالیسی اچھی ہے
nahi nawāz hakumath ki yeh pālisi achi hai
“No, this policy of Nawaz government is fine”.

(9) کچھ نہیں سب ڈرامے بازی ہے۔
kuch nahi sab dram-e bāzi hai
“It is nothing but a stage act”.

(10) ایسی بدعنوان حکومت نہیں ہونی چاہیے۔
‘aisi bad’ unwān hakumath nahi honi chāhiye
“A corrupt government has no right to rule”.

In examples (8), (9), and (10) negation words are present at two or more words distance from the underlined sentiment word. Therefore, the presence of these negation words does not have any effect on the polarity of these sentiment words. For example, if negation at two words difference is considered then in example (9) and (10), the underlined sentiment words polarity will be altered (from negative to positive) and hence will give wrong sentiment result.

However, examples (11), (12) and (13), taken from [4], show the presence of negation word at one word distance (before or after) from sentiment word. This occurrence of negation word affects the polarity of preceding or following sentiment words. Thus the

fetched polarity of sentiment word is altered by the classifier.

اُپریشن مسلے کا حل نہیں، اس سے بہت سے لوگ متاثر (11) ہونگے۔

Āpreshan masle ka hāl nahi, is se bohuth se log muthāsir hongy

“Military is not the solution as it adversely affects people”.

(12) سیاست دان قابل اعتماد نہیں
Siyāsathdān qābil –yi ‘aithmād nahi
“Politicians cannot be trusted”.

(13) بھارت کبھی پاکستان کا دوست نہیں ہو سکتا۔
Bahārat kabi Pākistān ka dost nahi ho sakthā
“India can never be a friend to Pakistan”.

5. Experimental Results

This section explains the experiment performed on corpus of Urdu opinionated data taken from the domain of news [4].

In the field of sentiment analysis, the performance of a system varies from one domain to another and similar is the case with the level of analysis selected. The performance of an algorithm that is good in one domain may not be the same when switched to another domain. [9] Performed sentiment analysis on two different corpora (product reviews corpus and movie reviews corpus). The same analyzer achieved 72% accuracy in the movie review domain while it achieved 78% accuracy in the product review domain. Therefore, there is no such agreed upon performance level in this field.

5.1. Experiment 1.

This experiment is performed to observe the opinions which have been classified as positive or negative on the basis of noun (NN) only, adjectives (ADJ) only or both nouns and adjectives together. Figure 2 shows a comparison chart of sentiment words (NN, ADJ, NN+ADJ), highlighting the importance of Nouns in the field of sentiment analysis. The corpus has been divided into sections each having 100 opinions. Each section is analyzed and figures are collected on the basis of following:

1. Opinions which contain only nouns
2. Opinions which contain only adjectives
3. Opinions which contain both nouns and adjectives

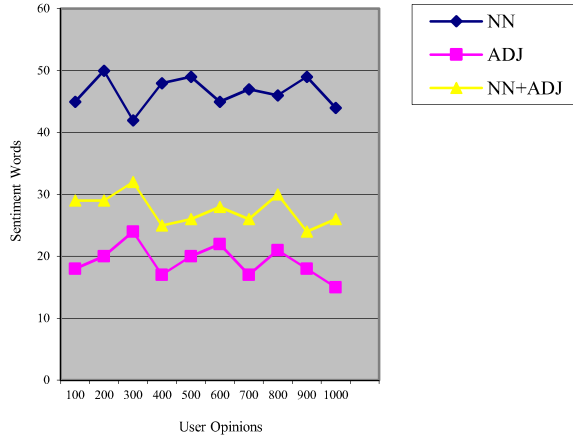


Figure 2: Sentiment Words Comparison

The experimental results show the usefulness of subjective nouns, where most of the decision is made by exploiting the nouns when there is lack of adjectives in data. However, nouns + adjectives also give better performance than only adjectives.

5.2. Experiment 2.

This experiment is performed to check the overall performance of the developed Urdu sentiment analyzer. A total of 1000 opinions have been analyzed consisting of positive, negative and neutral opinions from news domain. *Accuracy* is used as the system performance metric. It is the measure of how much close is the document classification suggested by our system to the actual sentiments present in the review. It is the percentage of correctly classified objects by the system, calculated by the following formula [32].

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

Where

- A is the accuracy of sentiment classifier.
- TP (True Positive) is the number of positive sentences that are correctly classified as positive.
- TN (True Negative) is the number of negative sentences that are correctly classified as negative.
- FP (False Positive) is the number of positive sentences that are incorrectly classified as positive.
- FN (False Negative) is the number of negative sentences that are incorrectly classified as negative.

The figures mentioned in Table 1 have been taken from the corpus exploiting the current Urdu sentiment analyzer:

Table 1: System Performance

TP	TN	FP	FN	Accuracy
448	420	62	70	86.8%

5.3. Analysis.

Considerable amount of work is done in the field of Urdu sentiment analysis. In [9] the analysis is performed on a corpus of movie and product reviews and achieved 74-76% of accuracy. [28] Extended their previous work mentioned in [9] by associating the targets with SentiUnits and improved the performance to 85%. The targets are the attribute features for which an opinion is made. The work reported in [29] adopted a lexicon-based approach for bilingual sentiment analysis of tweets using bilingual dataset (English and Roman Urdu) and achieved 76% accuracy. The current research, that uses the lexicon-based approach for Urdu opinionated data, has achieved a better state-of-the-art performance with 86.8% accuracy (Table 1). Due to the unavailability of prior Urdu opinionated data, the current classifier was tested on the corpus constructed in this research work.

6. Conclusion

Discussion in the previous sections clearly reflects that sentiment analysis is a growing field of Natural Language Processing which aims at developing automated tools for categorizing opinions as positive, negative or neutral that greatly help timely and effective decision making. There are also indications that sentiment analysis for Urdu language is in its infancy and needs to be explored. Though these are big challenges for young researchers but there are great opportunities too.

This study explains the process of developing Urdu sentiment analyzer by using nouns, in the domain of news, as sentiment carriers in addition to the traditional use of adjectives which are considered as universal sentiment carriers. It is proved that the importance of nouns as sentiment expressions cannot be denied especially in news domain and thus nouns help improve the performance of sentiment analyzer in the absence of adjectives.

•

7. Future Work

In this study, the scope of sentiment expressions has been extended to subjective nouns in the domain of news. However, in future other domains can be considered for the presence of nouns as sentiment expressions e.g. product and movie reviews. Similarly, experiments on different domain-POS combinations can be carried out.

The lexicon entries need to be expanded to cover maximum sentiment words of Urdu language in order to improve the performance of the system in future. Other parts of speech can also be examined and can be included in the lexicon. A synonym column can be employed which contains synonyms of sentiment words, so that if exact matching is not available then the decision can be made on the basis of appropriate matching with the help of synonym entry in the lexicon.

8. References

- [1] T. Nasukawa and J. Yi. "Sentiment analysis: Capturing favorability using natural language processing." In *pro. 2nd international conference on Knowledge capture*, ACM, 2003.
- [2] P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha. "Comparing and combining sentiment analysis methods," in *proc. of the first ACM conf. on Online social networks*, ACM, 2013.
- [3] X. Ding, B. Liu, and PS. Yu. "A holistic lexicon-based approach to opinion mining." In *proc. of the International Conference on Web Search and Data Mining*, ACM, 2008.
- [4] www.dailyjang.com, [access date 25 July 2014].
- [5] N. Durrani, and S. Hussain. "Urdu word segmentation." In *proc. of the Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 2010.
- [6] S. Hussain. "www. LICT4D. Asia/Fonts/Nafees_Nastalique." In *proc. 12th AMIC Annual Conference on E-Worlds: Governments, Business and Civil Society*, Asian Media Information Center, Singapore. 2003.
- [7] Y. Jo, and A. H. Oh. "Aspect and sentiment unification model for online review analysis." In *proc. fourth ACM international conference on Web search and data mining*, ACM, 2011.
- [8] T. T. Thet, J. C. Na, and C. S. G. Khoo. "Aspect-based sentiment analysis of movie reviews on discussion boards." *Journal of Information Science*. 2010.
- [9] A. Z. Syed, M. Aslam, A. M. Martinez-Enriquez, "Lexicon Based Sentiment Analysis of Urdu Text Using SentiUnits," In *proc. of the 9th Mexican Int. Conf. of Artificial intelligence*, Mexico, 2010.
- [10] A. Meena and T. V. Prabhakar. "Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis". Springer Berlin Heidelberg, 2007.
- [11] J. Zhao, K. Liu, and G. Wang. "Adding redundant features for CRFs-based sentence sentiment classification." In *proc. of the conference on empirical methods in natural language processing*, Association for Computational Linguistics, 2008.
- [12] G. Gezici, B. Yanikoglu, D. Tapucu, and Y. Saygin. "New features for sentiment analysis: Do sentences matter." In *SDAD 2012 The 1st international workshop on sentiment discovery from affective data*, p. 5. 2012.
- [13] K. Eguchi and V. Lavrenko. "Sentiment retrieval using generative models." In *proc. of the 2006 conference on empirical methods in natural language processing*, Association for Computational Linguistics, 2006.
- [14] B. Pang, L. Lee and S. Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." In *proc. of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, Association for Computational Linguistics, 2002.
- [15] P. D. Turney. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." In *proc. of the 40th annual meeting on association for computational linguistics*, Association for Computational Linguistics, 2002.
- [16] A. Khan. "Sentiment classification by sentence level semantic orientation using sentiwordnet from online reviews and Blogs." *International Journal of Computer Science & Emerging Technologies* 2, no. 4, 2011.
- [17] N. Farra, E. Challita, R. A. Assi, and H. Hajj. "Sentence-level and document-level sentiment mining for Arabic texts." In *IEEE International Conference on Data Mining Workshops (ICDMW), 2010*, IEEE, 2010.
- [18] S. Kim and E. Hovy. "Determining the sentiment of opinions." In *proc. of the 20th international conference on Computational Linguistics*, Association for Computational Linguistics, 2004.
- [19] R. McDonald, K. Hannan, T. Neylon, M. Wells, and J. Reynar. "Structured models for fine-to-coarse sentiment analysis." In *Annual Meeting-Association For Computational Linguistics*, vol. 45, no. 1, p. 432. 2007.
- [20] M. Hu, and B. Liu. "Mining and summarizing customer reviews." In *proc. of the tenth ACM SIGKDD international*

- conference on Knowledge discovery and data mining, pp. 168-177.ACM, 2004.
- [21] K. Denecke. "Are SentiWordNet scores suited for multi-domain sentiment classification?." In *Fourth International Conference on Digital Information Management, 2009. ICDIM*, pp. 1-6. IEEE, 2009.
- [22] B. Ohana and B. Tierney. "Sentiment classification of reviews using SentiWordNet." In *9th. IT & T Conference*, p. 13. 2009.
- [23] S. Mohammad, Cody Dunne, and Bonnie Dorr. "Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus." In *proc. of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, pp. 599-608. Association for Computational Linguistics, 2009.
- [24] S. Mukund, and R K. Srihari. "A vector space model for subjectivity classification in Urdu aided by co-training." In *proc. of the 23rd International Conference on Computational Linguistics: Posters*, pp. 860-868. Association for Computational Linguistics, 2010.
- [25] A. Z. Syed, M. Aslam, A. M. Martinez-Enriquez, "Sentiment analysis of Urdu language: Handling phrase level negation," In *proc. of the 10th int. conf. of Artificial intelligence*, Mexico, 2011.
- [26] A. Z. Syed, M. Aslam, A. M. Martinez-Enriquez, "Adjectival phrases as the sentiment carriers in Urdu text," *Journal of American Science*, vol. 7, no. 3, pp. 644-652, 2011.
- [27] S. Mukund and RK. Srihari. "Analyzing urdu social media for sentiments using transfer learning with controlled translations." In *proc. of the Second Workshop on Language in Social Media*, pp. 1-8. Association for Computational Linguistics, 2012.
- [28] A. Z. Syed, M. Aslam, A. M. Martinez-Enriquez, "Associating targets with SentiUnits: a step forward in sentiment analysis of Urdu text," *Artificial intelligence review*, pp. 1-27, March 2012.
- [29] I. Javed, H. afzal, A. Majeed and B. khan, "Towards Creation of Linguistic Resources for Bilingual Sentiment Analysis of Twitter Data", 2014.
- [30] B. Jawaaid, A. Kamran, O. Bojar, "A tagged corpus and a tagger for Urdu," In *proc. of the 9th Int. Conf. on language processing and evaluation*, Reykjavik, LREC 2014, Iceland, 2014.
- [31] B. Jawaaid and O. Bojar, "Tagger Voting for Urdu." In *24th International Conference on Computational Linguistics*, p. 135. 2012.
- [32] R. Prabowo and M. Thelwall. "Sentiment analysis: A combined approach." *Journal of Informetrics* 3, no. 2, p. 143-157, 2009.
- [33] Subrahmanian, V. S.; Reforgiato, D. "AVA: Adjective-verb-adverb combinations for sentiment analysis." *Intelligent Systems, IEEE* 23, no. 4, pp. 43-50, 2008.
- [34] Sokolova, M.; Lapalme, G. "A systematic analysis of performance measures for classification tasks." *Information Processing & Management* 45, no. 4, pp. 427-437, 2009
- [35] Hatzivassiloglou, V.; McKeown, K. R. "Towards the automatic identification of adjectival scales: Clustering adjectives according to meaning." In *Proc. of the 31st annual meeting on Association for Computational Linguistics*, Association for Computational Linguistics, 1993, pp. 172-182
- [36] Riloff, E.; Wiebe, J. "Learning extraction patterns for subjective expressions." In *Proc. of the 2003 Conf. on Empirical methods in natural language processing*, Association for Computational Linguistics, 2003, pp. 105-112.
- [37] Bloom, K.; Argamon, S. "Unsupervised extraction of appraisal expressions." In *Advances in Artificial Intelligence*, Springer Berlin Heidelberg, pp. 290-294, 2010.
- [38] Chesley, P.; Vincent, B.; Li Xu; Srihari, R. K. "Using verbs and adjectives to automatically classify blog sentiment", in *AAAI symposium on computational approaches to analysing weblogs*, AAAI-CAAW, pp. 27-29, 2006.
- [39] Benamara, F.; Cesarano, C. A.; Picariello; Recupero, D. R.; Subrahmanian V. S. "Sentiment analysis: Adjectives and adverbs are better than adjectives alone," In *Proc. of the int. Conf. on weblogs and social media*, ICWSM, 2007.