

# Hindi Subjective Lexicon : A Lexical Resource for Hindi Polarity Classification

Akshat Bakliwal, Piyush Arora, Vasudeva Varma

Search and Information Extraction Lab, LTRC,  
International Institute of Information Technology, Hyderabad, India.  
{akshat.bakliwal, piyush.arora}@research.iiit.ac.in, vv@iiit.ac.in

## Abstract

With recent developments in web technologies, percentage web content in Hindi is growing up at a lightning speed. This information can prove to be very useful for researchers, governments and organization to learn what's on public mind, to make sound decisions. In this paper, we present a graph based wordnet expansion method to generate a full (adjective and adverb) subjective lexicon. We used synonym and antonym relations to expand the initial seed lexicon. We show three different evaluation strategies to validate the lexicon. We achieve 70.4% agreement with human annotators and ~79% accuracy on product review classification. Main contribution of our work

- 1) Developing a lexicon of adjectives and adverbs with polarity scores using Hindi Wordnet.
- 2) Developing an annotated corpora of Hindi Product Reviews.

**Keywords:** Sentiment Analysis, Hindi Subjective Lexicon (HSL), Hindi Product Reviews, WordNet Traversal

## 1. Introduction

Hindi has 500+ million native speakers which are spread across the world and web pages<sup>1</sup> catering information in Hindi are also increasing. These web pages are important source of critical information. Individuals, governments and corporates can mine this information for useful insights. In the field of Sentiment Analysis, there is not much work done for Hindi and our work is among the first few works in this direction.

Adjectives and adverbs (Benamara et al., 2007) play an important role in expressing sentiments and opinions. In this paper, we present a method of building a subjective lexicon for Hindi. Our method depends only on one resource which is the WordNet. Using WordNet and Breadth First Graph traversal method, we construct the subjectivity lexicon. In our method, initially a small seedlist of words is decided along with their polarity. Using WordNet, this seedlist is populated based on the synonyms and antonyms of the words in the list. Here, we make an assumption that synonyms possess similar polarity and antonyms show opposite polarity. We validate the proposed approach for Hindi Language using three different strategies.

Main contribution of our work

1. Developing a lexicon of adjectives and adverbs with polarity scores using Hindi WordNet.
2. Developing an annotated corpora of Hindi Product Reviews.

Along with the lexicon generated using method proposed here, we generated lexicons using two other methods.

1. Bi-Lingual Dict: Using English Hindi wordnet linking (Arun Karthikeyan Karra, 2010) and English SentiWordNet (Esuli and Sebastiani, 2006; Baccianella et al., 2010), we generated a Hindi-SentiWordNet. For generating this lexicon, we followed the approach

mentioned in (Joshi et al., 2010). This lexicon is explained in details in *Section 4.3*.

2. Translated-Dict: Using Google translate<sup>2</sup> we translated English SentiWordNet word-by-word to Hindi. This process yielded a Hindi-SentiWordNet with 12086 Adjectives and 1509 Adverbs which were translated without any error.

In our opinion, these are a few limitations of the above mentioned approaches

- Bi-Lingual dictionaries may not account for all the words because of language variations. Words can be used in multiple context in either languages and context dependent word mapping is a tough task, error prone and requires manual efforts. Wordnet and its inherent relations provide much richer information to be mined while deriving a subjective lexicon for a particular language. We performed experiments with this lexicon (BiLing-Dict) and results are reported in *Table 5*.
- Using Translation<sup>3</sup> method for generating subjective lexicon, there is a high possibility of losing the context information and sometimes may have translation errors. We performed experiments with this lexicon (Translated-Dict) and results are reported in *Table 5*.

Rest of the paper is structured as follows: Section 2 presents a comprehensive view of the approach proposed in this research work. Section 3 gives details about the lexicon generated and description of product review dataset for Hindi language. In Section 4, we describe the different methods of evaluation along with results. In Section 5, we perform the analysis of this system. Related work in this area is discussed in Section 6. Section 7 presents the conclusions along with directions for our future work.

<sup>1</sup><http://bbc.co.uk/hindi>, <http://bhaskar.com>

<sup>2</sup><http://translate.google.com/>

<sup>3</sup>From rich resource to scarce resource language

## 2. Approach

Algorithm proposed in this research is dependent on a pre-annotated seedlist and WordNet. In the pre-annotated seedlist we have 45 adjectives and 75 adverbs. Composition of the seedlist is mentioned in *Table 1*. We extend the assumption made by (min Kim and Hovy, 2006) to antonyms. We assume synonym carries same sentiment/polarity and antonym show opposite sentiment/polarity as compared to the root word.

Adjectives	
Positive	15
Negative	15
Objective	15
Total	<b>45</b>
Adverbs	
Positive	25
Negative	25
Objective	25
Total	<b>75</b>

Table 1: Initial Seed List Composition

We make a hypothesis of traversing WordNet like a graph where words are connected to each other based on synonym or antonym relations. Consider each word in this list as a node of the graph. Each node has many in-links and many out-links. This is an undirected graph which is not fully connected i.e. not all the nodes are connected to every other node.

We perform Breadth First (BF) expansion of the seedlist. For every word in the seedlist, we identify its synonyms and antonyms and append with appropriate polarity in the seedlist. Just like BF traversal, we mark the expanded node as visited to avoid multiple time expansion of the same node.

Using the method explained here, we generate Adjective and Adverb subjective lexicon. Here, we copy the initial seedlist to a temporary seedlist and expand the temporary seedlist to final subjective lexicon. While performing BF traversal, we get (dequeue) a pair everytime from the top of the temporary seedlist. This pair contains word and its polarity. We check for this word if is already marked visited and exists in the final seedlist or not. If this word is marked or is in the final seedlist then we don't populate this word further, we just add the current polarity of this word to the polarity in the final list. But if this word is not marked visited or is not in the final list, we do three things

1. Add this word to the final list with the current polarity
2. Find out all the synonyms of this word and en-queue them in the temporary seedlist with the polarity same as the source word.
3. Find out all the antonyms of this word and en-queue them in temporary seedlist with opposite polarity. ( P -> N, O -> O, N -> P).

We continue this process till all the words in the temporary seedlist are explored or in other words till the temporary

seedlist becomes empty. When the temporary seedlist becomes empty the final seedlist contains adjectives (adverbs) and against each adjective (adverb) we have string of P's, N's and O's. Based on this we decide the final polarity of the word. Say for a word 'x' in the final seedlist we have string 's' made of P's, N's and O's, then

$$\begin{aligned}
 \text{Length of string } (s) &= Len \\
 \text{Number of P's in } s &= nP \\
 \text{Number of N's in } s &= nN \\
 \text{Number of O's in } s &= nO \\
 \text{Positive polarity of } x &= nP/Len \\
 \text{Negative polarity of } x &= nN/Len \\
 \text{Objective polarity of } x &= nO/Len
 \end{aligned} \tag{1}$$

For **Pseudo Code** Refer *Algorithm 1*.

## 3. Resources Created

The major contribution of this research are the two resources we have created. These resources are detailed in the following subsections (3.1 and 3.2).

### 3.1. Hindi Subjective Lexicon

Lexicon<sup>4</sup> built using the above mentioned approach for Hindi language contains 8048 adjectives and 888 adverbs. Hindi Subjective Lexicon is detailed in *Table 2*.

Adjectives	
Positive	2521
Negative	3060
Objective	2467
Total	<b>8048</b>
Adverbs	
Positive	186
Negative	175
Objective	527
Total	<b>888</b>

Table 2: Hindi Subjective Lexicon Details

### 3.2. Product Review Dataset

For product review dataset<sup>3</sup> in Hindi language we translated pre-annotated Amazon product reviews(Blitzer et al., 2007) from English to Hindi using Google translate<sup>5</sup>. All the translated reviews were of length  $\leq 25$ . We took this threshold of 25 words in order to avoid (reduce) translations errors. After translating the product reviews we asked human judges to manually validate (correct) the translation. *Table 3* summarize the data (reviews) generated by translation.

<sup>4</sup>This Resource is in the initial stage of development and is available for non-commercial and research usage on request. Request should be made to first or second author.

<sup>5</sup>We made an assumption that while translation sentiment bearing words are translated correctly without any loss or modification of sentiments.

---

**Algorithm 1** Algorithm for Populating SeedList using WordNet to generate Subjective Lexicon

---

```
Main() :  
  initialSeedList_Adj = 45 words (15 objective, 15 positive, 15 negative)  
  initialSeedList_Adv = 75 words (25 objective, 25 positive, 25 negative)  
  tempSeedList_Adj = initialSeedList_Adj  
  FinalDict_Adj = expand(tempSeedList_Adj)  
  
  tempSeedList_Adv = initialSeedList_Adv  
  FinalDict_Adv = expand(tempSeedList_Adv)
```

```
Expand(tempSeedList) :  
while tempSeedList  $\neq$  EmptyList do  
  Read the first pair from the tempSeedlist  
  pair = tempSeedList.dequeue()  
  // pair == < word, polarity >  
  if word  $\in$  FinalDict then  
    pol = FinalDict(word)  
    pol = pol + polarity  
    Finaldict(word) = pol  
    MarkVisited(word)  
    continue;  
  else  
    FinalDict(word) = polarity  
    synonymList = getSynonym(word) //Refer Algorithm 2  
    for all synonyms  $\in$  synonymList do  
      Add to tempSeedlist < synonym(i), polarity > at the bottom.  
    end for  
    antonymList = getAntonym(word) //Refer Algorithm 2  
    for all antonyms  $\in$  antonymList do  
      Add to tempSeedlist < antonym(i), opp - polarity > at the bottom.  
      // opp - polarity will be P if word has polarity N  
      // opp - polarity will be N if word has polarity P  
      // opp - polarity will be O if word has polarity O  
    end for  
  end if  
end while  
return FinalDict
```

---

Total Positive Reviews	1000
Manually Corrected Reviews	350
Total Negative Reviews	1000
Manually Corrected Reviews	350
Total Reviews	350 + 350

Table 3: Product Review Data Summary

## 4. Evaluation Methodologies and Results

Evaluation plays an important role while proposing a new approach to a problem. We mainly/commonly employ below mentioned strategies to evaluate.

1. Classification: In this method, we classify pre-annotated reviews using our system generated lexicon and find the accuracy to show the correctness. This strategy is generally applied to resource rich languages or for those languages for which we have pre-annotated data.

2. Validating Against Existing Resources: In this strategy, we find the accordance of lexicon generated using our approach with a lexicon which is already proposed and accepted by the research community.
3. Human Judgment: This method is usually opted for languages which are scarce resource languages. In this method, some manual annotators are appointed whose task is to annotate the lexicon generated and later, taking the majority vote of annotators the system generated lexicon is validated.

Subsequent sub-sections explain the three methods we used to evaluate the Hindi lexicon generated by our system.

### 4.1. Human Judgment

For this evaluation task, we hired five manual annotators<sup>6</sup> who are language experts in Hindi. We asked each annotator to tag the words generated by our system on the scale of

---

<sup>6</sup>None of the authors were annotators for this task.

---

**Algorithm 2** Algorithm for finding Synonyms and Antonyms of a word

---

```
getSynonym(word) :  
//every word in the wordnet is assigned a tag(s) based on the sense usage information if available.  
tag = getTag(word)  
synonymList = {}  
for all t ∈ tag do  
  words = getWords(t) // get all the words with this tag.  
  for all w ∈ words do  
    if w ∈ synonymList then  
      continue;  
    else  
      Add to synonymList at the bottom.  
    end if  
  end for  
end for  
return synonymList
```

```
getAntonym(word) :  
//every word in the wordnet is assigned a tag(s) based on the sense usage information if available.  
tag = getTag(word)  
antonymList = {}  
for all t ∈ tag do  
  words = getWords(t) // get all the words with this tag.  
  for all w ∈ words do  
    if w ∈ antonymList then  
      continue;  
    else  
      Add to antonymList at the bottom.  
    end if  
  end for  
end for  
return antonymList
```

---

3 (negative:-1, neutral:0, positive:1). After getting the list annotated by all the annotators, we had five votes for each word and we took the majority call. *Table 4* reports accordance of Hindi lexicon generated using our system with manual annotation. Reason behind low mutual agreement

Mutual agreement among the annotators	<b>70.48%</b>
Agreement of each annotator with our lexicon	
Annotator 1	66.08%
Annotator 2	64.01%
Annotator 3	<b>68.45%</b>
Annotator 4	66.70%
Annotator 5	68.34%
Overall Agreement of our lexicon with the annotators	<b>70.4%</b>

Table 4: Results for Manual Agreement for Hindi Lexicon

among the annotators is that many words in Hindi show ambiguous nature. Their polarity depends on the sense in which they are used. This ambiguous nature is highlighted in *Figure 1*.

## 4.2. Review Classification

For this evaluation strategy, we performed classification on product review dataset described in *Section 3.2*. On this data, we performed unigram presence and simple scoring method classification. We used a Shallow Parser<sup>7</sup> to identify adjectives and adverbs in a sentence. In unigram presence method, we count unigrams of positive, negative and objective polarity and assigned the polarity for which the count was highest. In simple scoring method, we summed the positive, negative and objective scores of each adjective (adverb) and assigned the polarity of the dominant score. From every review we identified adjective (adverb) and scored those adjectives (adverbs) using our lexicon. If an adjective (adverb) was missing from our lexicon we considered the stemmed variant<sup>8 9</sup> of that word for scoring. In addition to stemming we also performed negation handling. We identified the words with tag “NEG” (marked by Hindi Shallow Parser) and swapped the polarity (positive and negative) of adjectives (adverbs) in the sliding window

---

<sup>7</sup><http://ltrc.iit.ac.in/analyzer/hindi/shallow-parser-hin-3.0.fc8.tgz>

<sup>8</sup>We used the stemmer which is bundled with Hindi WordNet API 1.2

<sup>9</sup>[cflit.iitb.ac.in/wordnet/webhwn/index.php](http://cflit.iitb.ac.in/wordnet/webhwn/index.php)

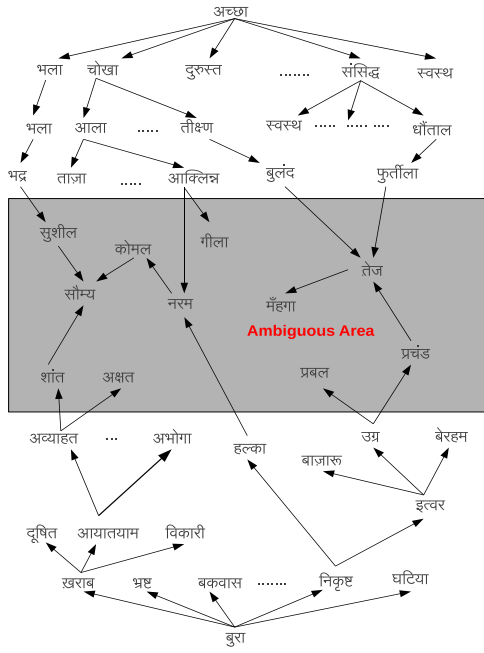


Figure 1: Figure shows the Graph Traversal of words in Hindi WordNet. The dark portion shows words which have ambiguous nature.

of 6 words. Our sliding window, looked upto 3 words in both the directions (left and right) of this word. Table 5 reports the results of classification.

### 4.3. Accordance with Existing Resources

(Joshi et al., 2010) proposed a method to generate a subjective lexicon for Hindi. Their method depends on two lexical resources: English SentiWordNet and English-Hindi WordNet Linking(Karthikeyan, 2010). Algorithm proposed in their research finds polarity of each Hindi word using SentiWordNet and English-Hindi BiLingual Mapping Dictionary. Using the mentioned approach, we generated the subjective lexicon. This lexicon contains 4335 adjectives and 1279 adverbs. One observation we made while generating this lexicon is that most of the adverbs are derived from adjective and nouns like luckily, happily, prettily, etc. Table 6 details number of adjectives and adverbs in each of the three lexicons.

## 5. Analysis

Results in Table 5 highlights the fact that our scoring method performs better than unigram presence method. Using this sample review “मोबाइल की गुणवत्ता अच्छी है लेकिन बैटरी जीवन भयानक है”, we explain how classification using our scoring method outperform the mere presence of unigrams. In this review, we have two adjectives (“अच्छी”:positive, “भयानक”:negative). When we account for unigram presence, the review is tagged as neutral but when we score using the scores in our lexicon, review is tagged as positive. Score for “अच्छी” in our lexicon is (1.0, 0.0, 0.0) and score for “भयानक” in our lexicon is (0.25, 0.75, 0.0). Overall score for this review becomes (+ 0.5) which correctly classifies the review.

Method	Accuracy (%)
<b>Adjective Presence</b>	
Bi-Lingual Dict (Baseline)	69.30
Translated-Dict (Baseline)	62.30
<i>Our Lexicon</i>	
Baseline	73.15
Baseline + NH	73.87
Baseline + Stem	75.70
Baseline + Stem + NH	76.74
<b>Adjective Scoring</b>	
Baseline	73.96
Baseline + NH	74.88
Baseline + Stem	77.44
Baseline + Stem + NH	<b>78.41</b>
<b>Adjective + Adverb Presence</b>	
Bi-Lingual Dict (Baseline)	71.40
Translated-Dict (Baseline)	63.90
<i>Our Lexicon</i>	
Baseline	73.70
Baseline + NH	73.89
Baseline + Stem	76.03
Baseline + Stem + NH	77.34
<b>Adjective + Adverb Scoring</b>	
Baseline	74.62
Baseline + NH	74.96
Baseline + Stem	78.27
Baseline + Stem + NH	<b>79.03</b>

Table 5: Results for Product Review Classification. (NH==Negation Handling)

	Our Lexicon	Bi-Ling Dict	Translated Dict
Adjective	8048	4335	12086
Adverbs	888	1279	1509

Table 6: Number of Words in each Lexicon

We proposed the use of stemmer to identify the root word for adjectives and adverbs which were present in the review but went missing from our lexicon. Stemming showed an improvement of ~3% in classification of reviews. Table 7 lists a few mapping of words to their stemmed form.

Word(s)	Stemmed Word
छोटे	छोटा
अच्छी, अच्छे	अच्छा
बड़ी	बड़ा
हल्के	हल्का
लंबे	लंबा

Table 7: Words and their stemmed (root) words

Along with stemming, we experimented with negation handling using the proposed strategy, results in Table 5 show ~3% improvement in classification of reviews. This section highlights the limitations and issues with the current version of algorithm proposed above.

- One of the major limitations of this system is that the current version of this algorithm does not perform Word Sense Disambiguation. This is so because Hindi WordNet lacks the information on most commonly used senses.
- Scope of the system proposed above is dependent on the initial seedlist used to populate the WordNet. If we choose the seed list in a careful manner with the help of linguistic experts the results and scope of the Lexicon thus generated would be better.
- For some adjectives and adverbs various morphological variants are used. Instead of using a stemmer if a morph is used in place then we expect results to improve. This is so because missing number of adjectives (adverbs) might reduce.

## 6. Previous Work

Many researchers have worked on various aspects of opinion analysis. (Pang et al., 2002), (Turney, 2002) worked on document level sentiment classification. (Wiebe et al., 1999), (Intelligent and Wilson, 2003), (Yu and Hatzivassiloglou, 2003), (min Kim, 2004), (Hu and Liu, 2004) worked on sentence level sentiment classification. More recently (Wilson, 2005), (Agarwal et al., 2009) worked at phrase level sentiment classification.

### 6.1. Non-Hindi Languages

Work on extracting sentiment from plain text was started way back in 1966 when IBM developed the General Inquirer System (Stone et al., 1966). IBM termed it as content analysis research problem in behavior science. Their system had 11789 total instances with each word having at least one instance. Hatzivassiloglou and McKeown (Hatzivassiloglou and McKeown, 1997) in 1997 started in the same direction of predicting the semantic orientation of adjectives. Their idea was predicting the semantic orientation of adjectives based on the nature of conjunctive joining the two adjectives. A log-linear regression model uses these constraints to predict whether conjoined adjectives are of same or different orientations, achieving 82% accuracy in this task when each conjunction is considered independently. In 2002 P.D. Turney (Turney, 2002) extended the work done by Hatzivassiloglou and McKeown. Turney used adverbs and nouns along with adjectives for performing opinion classification on reviews. He achieved 84% accuracy on automobile review classification compared to 66% on movie reviews.

(Kamps et al., 2004) try to determine sentiments of adjectives in WordNet. The polarity orientation of a word 'w' is measured as follows

$$O(w) = (d(w, good) - d(w, bad)) / d(good, bad) \quad (2)$$

where  $d()$  is word-net relatedness measure.  $O(t)$  belongs to range  $[-1, 1]$ , -1 for words on bad side and 1 for words on good side. Based on this method, they populated a total of 1608 words with avg. correctness of 67.18% for English. (Wiebe, 2000) showed a method to learn subjective adjectives from a corpora based on methods for clustering

words according to distributional similarity and a small amount of manually annotated corpora. (Carmen Banea and Wiebe, 2008) proposed a bootstrapping method for building subjective lexicon for under-resourced languages. Their method build a subjective lexicon using a small seedlist (60 words), an online dictionary (Romanian Dictionary) and a small annotated corpora.

For English, SentiWordNet (Baccianella et al., 2010; Esuli and Sebastiani, 2006) was developed in year 2006 by Esuli and Sebastiani. It contains four Part-of-Speech tags namely adjectives, adverbs, verbs and nouns with  $\sim 2$  million words out of which  $\sim 3\%$  are adjectives. Each word is assigned three scores positive, negative and objective which sum up to 1. SentiWordNet was build using WordNet and a ternary classifier. Their classifier is based on "bag of synset" model which uses manually disambiguated glosses available from the Princeton WordNet Gloss Corpus.

(min Kim and Hovy, 2006) proposed a method to classify a word as positive, negative or objective which was based on WordNet. They made an assumption which was to add synonyms of a word with the same polarity as the source word. To avoid words with multiple meaning (dual nature) they applied a method to identify closeness of a word to each category (positive, negative, objective). For their proposed method to give high recall the initial seedlist should be large enough and with wide variety words.

(Rao and Ravichandran, 2009) presented an extensive study on the problem of detecting polarity of words. They considered bi-polar classification of words i.e. a word can be either positive or negative. They performed semi-supervised label propagation in graph for polarity detection of words. Each of the words represent a node in the graph whose polarity is to be determined. They focused on three languages mainly English, French and Hindi but claim that their work can be extended to any other language for which WordNet is available.

(Das and Bandyopadhyay, 2010) developed sentiwordnet for Bengali language. They applied word level lexical-transfer technique to each entry English SentiWordNet using an English-Bengali Dictionary to obtain a Bengali SentiWordNet. This process resulted in 35,805 Bengali entries.

### 6.2. Hindi Language

(Joshi et al., 2010) created H-SWN (Hindi-SentiWordNet) using two lexical resources namely English SentiWordNet and English-Hindi WordNet Linking (Karthikeyan, 2010). Using WordNet linking they replaced words in English SentiWordNet with equivalent Hindi words to get H-SWN. (Arora et al., 2012) created Hindi Subjective lexicon using Hindi WordNet (Dipak Narayan and Bhattacharyya, 2002) as (only) resource. In (Arora et al., 2012), we built the lexicon for Adjectives only and tested the validity of our approach using two methodologies. This research is extension to the previous work (Arora et al., 2012), here we add Adverbs to the lexicon and also a new method (comparison with H-SWN) to validate the lexicon generated. As H-SWN was not available for public use, we built H-SWN using the method explained in (Joshi et al., 2010) and used this H-SWN for validation.

Our work is motivated towards Hindi Language and is re-

lated to works by Kim and Hovy (Kim and Hovy, 2006) and Rao and Ravichandran (Rao and Ravichandran, 2009). Kim and Hovy restricted their assumption to synonyms, we extend the relation to antonyms also. Rao and Ravichandran performed bi-polar classification, we extend classification to third level i.e. objectivity. In this work, we use Hindi WordNet (Dipak Narayan and Bhattacharyya, 2002) to obtain the polarity of adjectives and adverbs for Hindi Subjective Lexicon.

## 7. Conclusion and Future Work

Sentiment Analysis for Hindi is an important task. In this paper, we proposed a graph based method to generate the Hindi subjectivity lexicon. We explored how the synonym and antonym relations can be exploited using simple graph traversal to generate the subjectivity lexicon. We have tested and verified this approach for Hindi, but we believe this approach will work for any language. This approach just uses only one resource (WordNet) for Lexicon generation. Our proposed algorithm achieved  $\sim 79\%$  accuracy on classification of reviews and 70.4% agreement with human annotators.

In future, this work can be extended to incorporate Word Sense Disambiguation (WSD) and morphological variants which could result in better accuracy for words which have dual nature. We experimented with adjectives and adverbs, this work can be extended for other parts of speech (verbs and nouns).

## Acknowledgements

We would like to thank Dr. Dipti Mishra Sharma and her team for helping us with manual annotation task. We would also like to extend our thanks to Kushal Dave and Naveen kulkarni for their useful insights while writing the manuscript.

## 8. References

- Apoorv Agarwal, Fadi Biadry, and Kathleen R. McKeown. 2009. Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams.
- Piyush Arora, Akshat Bakliwal, and Vasudeva Varma. 2012. Hindi subjective lexicon generation using wordnet graph traversal.
- Rohan Raikar, Aditya Sharma, Pushpak Bhattacharyya, Arun Karthikeyan, Prabhakar Pande. 2010. Hindi english wordnet linkage. In *Dual degree thesis, IIT Bombay*.
- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*.
- Farah Benamara, Carmine Cesarano, Antonio Picariello, Diego Reforgiato, and V. S. Subrahmanian. 2007. Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*.
- John Blitzer, Mark Dredze, and Fernando Pereira. 2007. Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification. In *ACL*.
- Rada Mihalcea, Carmen Banea, and Janyce Wiebe. 2008. A bootstrapping method for building subjectivity lexicons for languages with scarce resources. In Bente Maegaard, Joseph Mariani, Jan Odjik, Stelios Piperidis, Daniel Tapias, Nicoletta Calzolari (Conference Chair), Khalid Choukri, editor, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May. European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2008/>.
- Amitava Das and Sivaji Bandyopadhyay. 2010. *SentiWordNet for Bangla*.
- Prabhakar Pande, Dipak Narayan, Debasri Chakrabarti, and P. Bhattacharyya. 2002. An experience in building the indo wordnet - a wordnet for hindi. In *First International Conference on Global WordNet*.
- Andrea Esuli and Fabrizio Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC06)*.
- Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, ACL '98*, pages 174–181, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *KDD*, pages 168–177.
- Theresa Wilson, Intelligent, and Theresa Wilson. 2003. Annotating opinions in the world press. In *In SIGdial-03*, pages 13–22.
- Aditya Joshi, Balamurali A R, and Pushpak Bhattacharyya. 2010. A fall-back strategy for sentiment analysis in hindi: a case study.
- Jaap Kamps, Maarten Marx, Robert J. Mokken, and Maarten De Rijke. 2004. Using wordnet to measure semantic orientation of adjectives. In *National Institute for*, pages 1115–1118.
- Arun Karthikeyan. 2010. Hindi english wordnet linkage.
- Soo min Kim and Eduard Hovy. 2006. Identifying and analyzing judgment opinions. In *Proceedings of HLT/NAACL-2006*, pages 200–207.
- Soo min Kim. 2004. Determining the sentiment of opinions. In *In Proceedings of COLING*, pages 1367–1373.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 79–86.
- Delip Rao and Deepak Ravichandran. 2009. Semi-supervised polarity lexicon induction. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, EACL '09*, pages 675–682, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, and

- Daniel M. Ogilvie. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.
- Peter Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews.
- Janyce M. Wiebe, Rebecca F. Bruce, and Thomas P. O'Hara. 1999. Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, ACL '99, pages 246–253, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Janyce Wiebe. 2000. Learning subjective adjectives from corpora. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*. AAAI Press.
- Theresa Wilson. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *In Proceedings of HLT-EMNLP*, pages 347–354.
- Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, EMNLP '03, pages 129–136, Stroudsburg, PA, USA. Association for Computational Linguistics.