

# HSAS: Hindi Subjectivity Analysis System

Vandana Jha\*, Manjunath N\*, P Deepa Shenoy\* and Venugopal K R\*

\*Department of Computer Science and Engineering

University Visvesvaraya College of Engineering, Bangalore University, Bangalore, India

Email: vjvandanajha@gmail.com

**Abstract**—With the development of Web 2.0, we are abundant with the documents expressing user’s opinions, attitudes and sentiments in the textual form. This user generated textual content is an important source of information to make sound decisions by the organizations and the government. The textual information can be categorized into two types: facts and opinions. Subjectivity analysis is the automatic extraction of subjective information from the opinions posted by users and divides the content into subjective and objective sentences. Most of the works in subjectivity analysis exists for English language data but with the introduction of unicode standards UTF-8, Hindi language content on the web is growing very rapidly. In this paper, Hindi Subjectivity Analysis System (HSAS) is proposed. It explores two different methods of generating subjectivity lexicon using the available resources in English language and their comparative evaluation in performing the task of subjectivity analysis at the sentence level. The first method uses English language *OpinionFinder* subjectivity lexicon. The second method uses a small seed word list of Hindi language and expands it to generate subjectivity lexicon. Different evaluation strategies are used to validate the lexicon. We achieved 71.4% agreement with human annotators and ~80% accuracy in classification on a parallel data set in English and Hindi. Extensive simulations conducted on the test dataset confirm the validity of the suggested method.

**Keywords**—Data Mining, Text Mining, Subjectivity Analysis, Hindi Language, Natural Language Processing

## I. INTRODUCTION

The textual information on the World Wide Web can be categorized into two main types: facts and opinions [1]. Facts are the objective statements whereas opinions are subjective statements. Human emotions and feelings towards entities, events and their properties are captured in such subjective statements. With the increasing popularity of the Web 2.0, it is very easy to post our opinion on any topics in Web forums and similar digital platforms. It is equally easy to access the information posted by others on the web by the concerned organizations or the government to make sound decisions for the improvement in their products, services or policies. Most of the time, these opinionated information are hard to detect in long reviews, blogs and posts and it is an uphill task for a person to separate the subjective opinion from the objective facts. Hence there is a need to perform subjectivity analysis which can do the task automatically.

Subjectivity Analysis is a fast developing sub-area of Natural Language Processing (NLP) and Text Mining. It is the extraction of subjective information from the input text and classifying a sentence into either objective type or subjective type [2]. The subjective information [3], [4] are the words

or a group of words used to express personal feelings such as sentiments, opinions, emotions, assessments, faith, and guesswork in natural language. For example, below is the sample where subjective and objective sentences are combined in a news article and subjective sentence is shown in bold:

*The petrol prices have gone down by Rs. 2 and that made general public **erupt with joy.***

Here, first part of the sentence is objective and second part is subjective.

The task of sentiment analysis can be divided into two sub problems. First, dividing the opinion posts into subjective and objective sentences and Second, dividing the subjective sentences into positive, negative and neutral classes. The first sub-problem of sentiment analysis is known as subjectivity analysis and the paper addresses this part of the sentiment analysis.

Sometimes subjectivity analysis becomes more challenging than polarity classification [5], [6]. The subjectivity of words and idioms is also context dependent and an objective script can have subjective sentences in it (e.g., people’s opinions can be quoted in a news article). Moreover, as said by Su, [7] when annotating texts, results are mostly subjectivity dependent. However, Pang [8] showed that separating objective statements from a script before analysing its polarity assisted in improving performance. Thus, this part of the sentiment analysis is worth exploring. Subjectivity Analysis can be performed at three levels: Document level, Sentence level and Word or Phrase level. There are three main directions that have been considered for all these levels of subjectivity annotations: (1) manual annotations, which involve human judgement of selected words and phrases; (2) automatic annotations, which involve knowledge sources such as dictionaries; and (3) automatic annotations, which involve information derived from corpora. Many works in the area of subjectivity analysis depends on the subjective meaning of group of words and phrases, also known as, subjectivity lexica. Our work is also using subjectivity lexicon for subjectivity analysis and addresses the problem of dividing Hindi language user-generated textual information into subjective and objective sentences.

### A. Motivation

Most of the research works in subjectivity analysis have done on English data. For a new language, the cost and efforts required to create corpora and tools have restricted the growth of subjectivity analysis in non-English languages. Despite this, work in other languages is increasing. It is essential to target non-English languages for subjectivity analysis as well, because only 28.6 percent of browsers are well-versed in English. We are performing subjectivity analysis in Hindi language. Hindi, the official language of India, has 490 million speakers

across the world, which is 4.7% of the world population and is the 4th largest spoken language in the world [9]. With the introduction of UTF-8 standards, Web pages in Hindi language is increasing very fast. But it is a challenging task for a resource scarce language like Hindi. This milestone can be achieved by using the means and mechanisms available in English, thus bypassing the overhead incurred in generating the resources in Hindi. As a part of research, we analyse ways for developing subjectivity analysis resources in Hindi using the means and mechanisms available in English and attempt to find possible solutions:

1. A subjectivity lexicon of high-quality in Hindi language can be derived from an available subjectivity lexicon in English language by using machine translation method.
2. A small seed list of subjectivity lexicon in Hindi language can be expanded and this maintains the quality of the lexicons in Hindi language.

### B. Contribution

In this paper, HSAS is proposed. It investigates two different methods of generating subjectivity annotated lexicon in Hindi language. First, is using a subjectivity lexicon available in English language and automatically translating it into Hindi language by using a bi-lingual dictionary. The resources required for this are a subjectivity annotated lexicon in English language and a bi-lingual dictionary. Second, is employing a modest seed list of subjectivity lexicon in Hindi language and growing it into subjectivity lexicon of comparable size by using resources available in Hindi language. The resources required for this are a selective small list of subjective word seeds in Hindi language, a bi-lingual dictionary and a raw corpus. We run simulations on Hindi language data to show the adaptability of the investigated methods.

### C. English Language Resources

This work uses the following English language resource to perform subjectivity analysis in Hindi language:

*OpinionFinder Subjectivity Lexicon.* An English language subjectivity lexicon which is used here is present in OpinionFinder [10], a subjectivity analysis system which naturally illustrates the subjectivity of a new text based on the existence (or non-existence) of words or phrases in a large lexicon. The lexicon contains 8227 entries of length 1 word, that is multi word expressions are avoided. These words are either of *strong* subjective type or of *weak* subjective type. The words whose presence change the context of full sentence into subjective are considered as strong indication of subjectivity whereas the words which have lesser presence, but still have more frequency, in subjective context are considered as weak clues of subjectivity. Each word is also labelled with a polarity tag, indicating whether it is positive, negative, or neutral. Let's say, we have one entry from the *OpinionFinder* lexicon *type=weaksubj len=1 word1=abandon pos1=verb stemmed1=y polannsrc=tw mpqapolarity=strongneg*, which indicates that the word *abandon* which is a verb is a weak indication of subjectivity and has a polarity that is strongly negative. As illustrated, each entry is also extended with the extra information, such as length, whether stemming is

performed on the word or not (stemmed=y or n), source, and hence forth.

*hindencorp.* This is a parallel corpus in English and Hindi language [11]. This corpus covers several areas in politics, sports, education, fashion and others and balanced in nature. We have selected a subset of this corpus containing 501 sentences in English and its manual translation present in the corpus in Hindi. This subset will act as test dataset for our model and satisfies the need of comparative evaluations.

### D. Organization

The paper is presented as follows: A brief overview of the related work is provided in section II. Section III describes two methods of generating subjectivity lexicon in Hindi language and evaluates their quality. Experiments run on Hindi language data to demonstrate the portability of the described methods is given in section IV. Section V concludes the paper.

## II. RELATED WORK

Most researchers have concentrated on sentiment analysis because of the demanding task of investigating product or movie reviews (again prompted by the presence of huge amount of online data in different languages which can be used here in subjectivity analysis), we refer works on subjectivity as well as polarity, because of their comparable methods, though the final outputs are not comparable.

### Non-Indian languages Research

Sentiment analysis research started in 1966 when the General Inquirer system was developed by IBM [12]. IBM termed it as content analysis research problem in behaviour science and contains 11789 words and each having at-least one instance. In 1998, [13] gave a method to anticipate semantic orientations of adjectives. They predicted the semantic orientations of adjectives used in conjunctions, that is, in a sentence, <adjective> and <adjective>, the adjectives must be of the same polarity. They achieved 82% accuracy. This finding was further taken up by Turney [14] to other POS-tags in 2002. They used five patterns for performing polarity classification on reviews. They found out that automobile review data had 84% accuracy while movie reviews had 66%. A high percentage of work in subjectivity analysis depends on generating subjective lexicon. E. sulis and Sebastiani developed SentiWordNet [15], [16] in year 2006. It has about 2 million words of four Part-of-Speech tags namely adjectives, adverbs, verbs and nouns. Every word is attached with positive, negative and objective scores with total equals to 1. WordNet and a ternary classifier were used to build SentiWordNet.

Banea et. al. [17] created a subjective lexicon in Romanian language by bootstrapping method using a selected small list of 60 words, an online dictionary and a small annotated corpus. LSA, a word level similarity is used to filter words. Kamps et. al. [18] developed a distance measure on WordNet, and showed how it can be utilized to figure out the semantic orientation of adjectives. With an approximation of 67.18% for English, they populated total 1608 words in all four categories. Paper [19] gave a method of determining/analysing judgement opinions in a four step process. The first step, identifying the opinion; the second step, identifying the valence; the third step, identifying the holder and last step, identifying the topic.

Rao and Ravichandran [20] gave a comprehensive study on the complication of recognizing word polarity. They considered a word can be classified bipolarity. They treated polarity detection as a semi-supervised label propagation problem in a graph. To determine the polarity of a word, it was represented by a node in the graph and to represent a relationship between two words, they were encoded by a weighted edge. Each node (word) can have two labels: positive or negative. They concentrated on languages English, French and Hindi but affirmed that the same methodology can be applied to any language which has WordNet. Association rule mining using Genetic Algorithm is used in the papers [21], [22], [23].

#### Indian languages Research

Comparatively less research has been done for Indian languages. Paper [24], suggested a computational technique for developing Senti-WordNet(Bengali) using English-Bengali bilingual dictionary and English Sentiment Lexicons. They successfully got 35,805 Bengali words by applying lexical-transfer technique at word level to each word in English SentiWordNet using an English-Bengali Dictionary to obtain a Bengali SentiWordNet. Das and Bandopadhyaya [25], introduced four approaches to predict the polarity of a word. An interactive game is provided to identify the polarity of words in first strategy. In second strategy, a bilingual dictionary is developed for English and Indian Languages. In third strategy, word net expansion is done using antonym and synonym relations. In fourth approach, a pre-annotated corpus is used for learning. Paper [26], developed the method for tagging using the Bengali words. Classification of words is performed into six emotion classes according to three categories of intensities (low, general and high).

In paper [27], Hindi Subjective Lexicon and hindi WordNet has been used for the identification of semantic orientation of adjectives and adverbs. By using a graph based method Bakliwal et al. [28] created subjectivity lexicon. Namita Mittal et al. [29] developed an efficient approach based on negation handling and discourse relation to identify the sentiments from Hindi content. They developed an annotated corpus for Hindi language and improved the existing Hindi SentiWordNet (HSWN) by including more opinion words into it. Their proposed algorithm achieved approximately 80% accuracy on classification of reviews. Paper [30] proposed a stopword removal algorithm for Hindi Language which is based on a Deterministic Finite Automata (DFA). They achieved 99% accurate results. Paper [31] proposed a multi-domain sentiment aware dictionary.

### III. PROPOSED WORK

Manually or semi-automatically constructed lexicons are the starting point for subjectivity and sentiment analysis [19], [4], [32]. In this research work, the first method is based on creating a subjectivity lexicon in Hindi language using English language resource and detailed in subsection A. The second method is expanding the selected seed-list of subjective lexicon in Hindi language and creating its own language specific subjectivity lexicon. It is detailed in subsection B. Both the approaches have been evaluated by constructing a rule-based classifier to classify the sentences into subjective and objective on the parallel dataset in English and Hindi language and their comparative evaluations.

#### A. Creating Subjectivity Lexicon

The most common approach for creating subjectivity lexicon in Hindi language is the translation of an existing English language lexicon by using a dictionary or translator. Here, we construct a subjectivity lexicon for Hindi by starting with the English subjectivity lexicon from *OpinionFinder* and translating it using translator<sup>1</sup> as well as English-Hindi bilingual online dictionary<sup>2</sup>. The translation process is a challenging task. The English subjectivity lexicon consists of inflected words. The words whose inflected form is available in either translator<sup>1</sup> or in dictionary<sup>2</sup>, those words are translated as it is. Only for the words whose inflected form translation is not available, they are first lemmatized and then translated. The words when lemmatized may lose its subjective meaning. Table I shows a sample from Hindi lexicon along with their English original form. The steps for this purpose is given in Algorithm

---

#### Algorithm 1: Subjective or Objective Classification

---

**Data:** Input Data file and OpinionFinder dictionary  
**Result:** Sentences labelled as Subjective or Objective and their count

```

begin
  Initialize:
  strong_subj_words_count = 0;
  weak_subj_words_count = 0;
  strong_subjective = [];
  weak_subjective = [];
  objective = True;
  Perform:
  Parse each sentence from the input file
  for each word in sentence do
    if word in dictionary then
      if wordtype in dictionary is strongsubj then
        strong_subj_words_count += 1
        if strong_subj_words_count > 0 then
          objective = false
        end
      else
        if wordtype in dictionary is weaksubj
        then
          weak_subj_words_count += 1
          if weak_subj_words_count > 1 then
            objective = false
          end
        end
      end
    end
  end
end
return objective
end

```

---

1 and its principle is as follows: The English (Hindi) input file is first parsed at the sentence level and for each sentence, it is parsed at word level. When the word is matched with the word present in the English (Hindi translated) *OpinionFinder* dictionary then its *word\_type* is checked. If it is strong subjective type then its *strong\_subj\_words\_count* is maintained.

<sup>1</sup><https://translate.google.co.in/>

<sup>2</sup><http://www.shabdikosh.com/>

TABLE I: A Sample of Hindi Subjectivity Lexica

English Word	Associated attributes	Hindi Word
luck	strongsubj, noun, positive	भाग्य
renunciation	strongsubj, noun, negative	संन्यास
bankrupt	weaksubj, adj, negative	दिवालिया
exclusively	weaksubj, adj, neutral	केवल
loot	strongsubj, verb, negative	लूटना
understand	strongsubj, verb, positive	जानना

TABLE II: A Sample of the Seed Words

Noun (translation in English)	संन्यास (renunciation), कष्ट (torment), भाग्य (luck), शांति (peacefulness), उदारता (nobleness), इच्छा (desirability)
Adjective (translation in English)	भयकर (fearful), खराब (horrible), बंजर (barren), असभ्य (barbaric), दिवालिया (bankrupt), तुच्छ (scant)
Adverb (translation in English)	लाभप्रद (gainfully), पूर्णतया (absolutely), केवल(exclusively), उत्साह(zealously), परिहासपूर्वक (facetiously), काफी (considerably)
Verb (translation in English)	आरोप (impeach), निश्चयपूर्वक (allege), लूटना (loot), जानना (understand), बढ़ाना (multiply), सूखना (dwindle)

Similarly *weak\_subj\_words\_count* is also maintained. If one strong subjective word occurs then the sentence is labelled as subjective sentence. For weak subjective words, sentences are labelled as subjective if its occurrence is two. In results, we have shown results for weak subjective occurrence of two as well as three.

### B. Subjectivity Lexicon Expansion

Initially, 60 seed words are selected randomly from the translated *OpinionFinder* dictionary. Here, Noun, Verb, Adjective and Adverbs categories are considered and each type has 15 words. Thus, the primary seed list is balanced and helps in good coverage of each part of speech category. Table II shows an example of the seed words. Algorithm 2 is for Lexicon expansion and its principle is as follows: For each seed word, all related words are collected from publicly available Hindi Wordnet [33]. These related words are synonyms, antonyms or any other word present in the definition of the seed word. We calculated the Pointwise Mutual Information (PMI) score for filtering some of the related words whose score is zero that means these are not occurring in the corpus. For calculating PMI score, we have collected a large corpus of Hindi files. The statistics for this is given in table III. After each new word is added in the initial seed word list, the above explained process is repeated to collect more subjective words. At the end, we have a collection of 4320 Hindi language subjective words which acts as dictionary. These words are used to find out subjective sentences from the Hindi data set (from the parallel data set) and performed slightly better than the translated dictionary results.

<sup>3</sup><http://www.hindinovels.net/>

TABLE III: Hindi Corpus Statistics

Type of File	Number	Size	Sentences	Words
Hindi Wikipedia [34]	506 files	42.8 MB	~700 sentences per file	~10 words per sentence
Hindi Novels <sup>3</sup>	10 files	9 MB	~1430 sentences per file	~10 words per sentence
Hindi Movie Reviews [35]	200 files	1.6 MB	~50 sentences per file	~8 words per sentence

### Algorithm 2: Subjectivity Lexicon Expansion

**Data:** Initial Seed Word List

**Result:** Expanded Seed Word List

**begin**

**Initialize:**

*words* = [];

*related\_words* = [];

*bi\_gram\_word\_formed* = "";

*final\_bi\_gram* = [];

*Dictionary* = *Dict*();

**Perform:**

Parse the *seed\_word* and *related\_words*

**for each line in a document do**

word.append(*seed\_word*)

related\_word.append(*related\_word*)

*Dictionary*[*seed\_word*] = *related\_word*

**end**

**for each word in words do**

**for each related\_word in related\_words do**

*bi\_gram\_word\_formed* += word +

*related\_word*

*final\_bi\_gram*.append(*bi\_gram\_word\_formed*)

**end**

**end**

Calculate PMI score with the generated *bi\_grams* and add the words with PMI score > 0

**end**

## IV. EXPERIMENTAL RESULTS

The results obtained by translating the English language resource *OpinionFinder* subjectivity lexicon into Hindi language to create Hindi subjectivity lexicon and evaluating it against the parallel dataset is shown in table IV. The result where every one occurrence of strong subjective word and two occurrence of weak subjective word are considered for labelling a sentence as subjective is given in Fig. 1. The results where every one occurrence of strong subjective word and three occurrence of weak subjective word are considered for labelling a sentence as subjective is given in Fig. 2. Fig. 3 displays the results for English Input file with *OpinionFinder* dictionary with one occurrence of strong subjective count and both (two as well as three) occurrences of weak subjective counts. Fig. 4 displays the results for Hindi Input file with Hindi translated *OpinionFinder* dictionary with one occurrence of strong subjective count and both (two as well as three) occurrences of weak subjective counts. The results are given to two human experts, native Hindi speaker, for subjectivity

TABLE IV: Results of Classification on Parallel Dataset

Language		Features								Total
		Without Preprocessing		With Stopword removal(SWR)		With Stemming(ST)		With SWR + ST		
		>1	>2	>1	>2	>1	>2	>1	>2	
English	Subj.	38	38	239	199	328	281	320	275	501
	Obj.	463	463	262	302	173	220	181	226	
Hindi	Subj.	349	231	292	185	283	162	237	127	501
	Obj.	152	270	209	316	218	339	264	374	

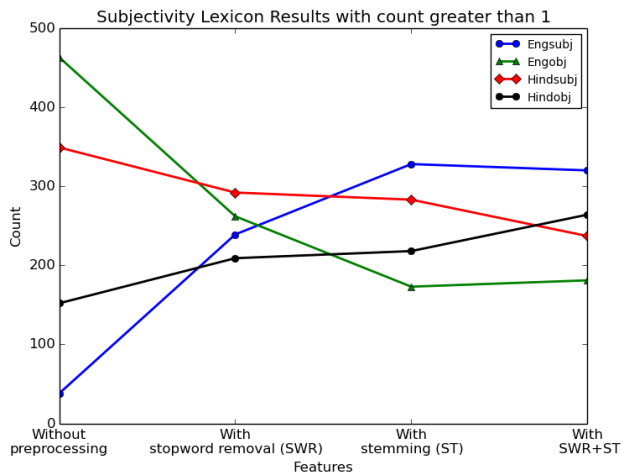


Fig. 1: Results when two weak subjective words are considered for labelling sentence as subjective.

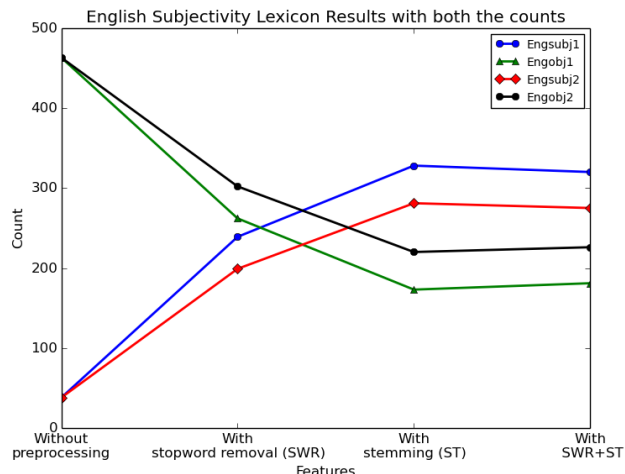


Fig. 3: Results of English Subjectivity Analysis

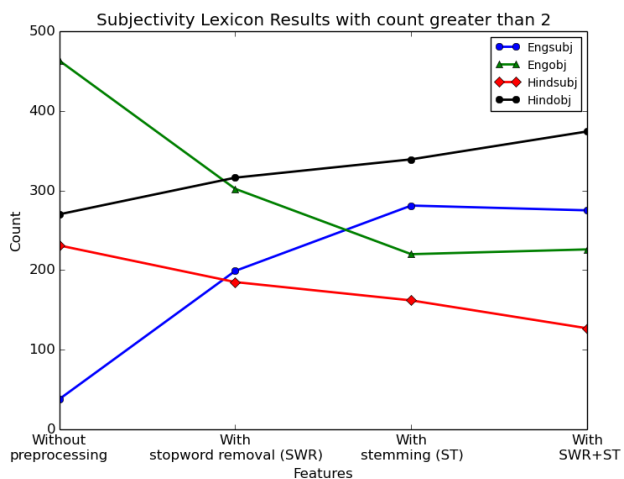


Fig. 2: Results when three weak subjective words are considered for labelling sentence as subjective.

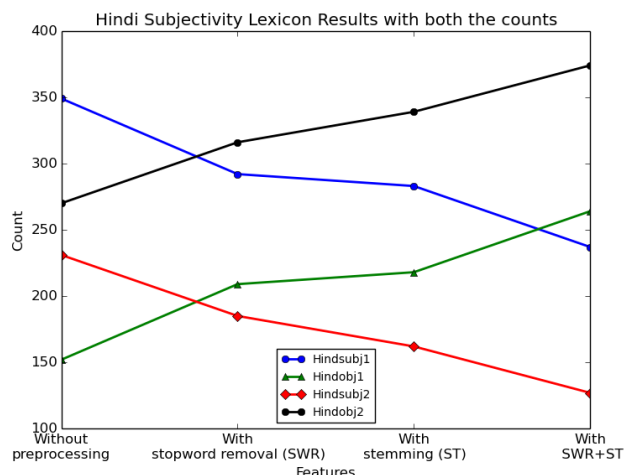


Fig. 4: Results of Hindi Subjectivity Analysis

## V. CONCLUSIONS

classification. With 71.4% agreement with human annotators, the results are ~79% accurate in classification. When Hindi dictionary (Expanded Seed word Lexicon) is used for the classification then the accuracy is slightly improved to ~80%.

With the development of Web 2.0 and introduction of UTF-8 unicode standards, the contents in Hindi language is available in abundance on the World Wide Web. We have proposed Hindi Subjectivity Analysis System (HSAS) in this paper. It gives two different methods of generating subjectivity lexicon and classifying the statements into subjective and

objective statements. The first method uses English language *OpinionFinder* subjectivity lexicon. The second method uses 60 seed words of Hindi language and expands the seed list to generate 4320 words subjectivity lexicon. We achieved 71.4% agreement with human annotators and ~80% accuracy in classification on a parallel data set in English and Hindi. In future, we are planning to expand the lexicon further. Moreover, for PMI score calculation also, we are enlarging the Hindi corpus files.

## REFERENCES

- [1] B. Liu, "Sentiment analysis and subjectivity," *Handbook of natural language processing*, vol. 2, pp. 627–666, 2010.
- [2] B. Pang and L. Lee, "4.1.2 subjectivity detection and opinion identification," *Opinion mining and sentiment analysis*, 2008.
- [3] J. Wiebe, T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language," *Language resources and evaluation*, vol. 39, no. 2-3, pp. 165–210, 2005.
- [4] E. Riloff, J. Wiebe, and T. Wilson, "Learning subjective nouns using extraction pattern bootstrapping," in *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*. Association for Computational Linguistics, 2003, pp. 25–32.
- [5] R. Mihalcea, C. Banea, and J. M. Wiebe, "Learning multilingual subjective language via cross-lingual projections," 2007.
- [6] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, 2005, pp. 347–354.
- [7] F. Su and K. Markert, "From words to senses: a case study of subjectivity recognition," in *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2008, pp. 825–832.
- [8] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2004, p. 271.
- [9] [Online]. Available: [http://en.wikipedia.org/wiki/List\\_of\\_languages\\_by\\_number\\_of\\_native\\_speakers](http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers)
- [10] T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan, "Opinionfinder: A system for subjectivity analysis," in *Proceedings of hlt/emnlp on interactive demonstrations*. Association for Computational Linguistics, 2005, pp. 34–35.
- [11] O. Bojar, V. Diatka, P. Rychlý, P. Straňák, V. Suchomel, A. Tamchyna, and D. Zeman, "HindEnCorp - Hindi-English and Hindi-only Corpus for Machine Translation," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, N. C. C. Chair, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, Eds. Reykjavik, Iceland: European Language Resources Association (ELRA), may 2014.
- [12] P. J. Stone, D. C. Dunphy, and M. S. Smith, "The general inquirer: A computer approach to content analysis." 1966.
- [13] V. Hatzivassiloglou and K. R. McKeown, "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*. Association for Computational Linguistics, 1997, pp. 174–181.
- [14] P. D. Turney, "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*. Association for Computational Linguistics, 2002, pp. 417–424.
- [15] A. Esuli and F. Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining," in *Proceedings of LREC*, vol. 6. Citeseer, 2006, pp. 417–422.
- [16] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining." in *LREC*, vol. 10, 2010, pp. 2200–2204.
- [17] C. Banea, J. M. Wiebe, and R. Mihalcea, "A bootstrapping method for building subjectivity lexicons for languages with scarce resources," 2008.
- [18] J. Kamps, M. Marx, R. J. Mokken, and M. De Rijke, "Using wordnet to measure semantic orientations of adjectives." in *LREC*, vol. 4. Citeseer, 2004, pp. 1115–1118.
- [19] S.-M. Kim and E. Hovy, "Identifying and analyzing judgment opinions," in *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*. Association for Computational Linguistics, 2006, pp. 200–207.
- [20] D. Rao and D. Ravichandran, "Semi-supervised polarity lexicon induction," in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2009, pp. 675–682.
- [21] P. D. Shenoy, K. Srinivasa, K. Venugopal, and L. M. Patnaik, "Evolutionary approach for mining association rules on dynamic databases," in *Advances in knowledge discovery and data mining*. Springer, 2003, pp. 325–336.
- [22] P. D. Shenoy, K. Srinivasa, K. Venugopal, and L. M. Patnaik, "Dynamic association rule mining using genetic algorithms," *Intelligent Data Analysis*, vol. 9, no. 5, pp. 439–453, 2005.
- [23] V. H. Bhat, P. G. Rao, R. Abhilash, P. D. Shenoy, K. Venugopal, and L. Patnaik, "A data mining approach for data generation and analysis for digital forensic application," *IACSIT International Journal of Engineering and Technology*, vol. 2, no. 3, pp. 314–319, 2010.
- [24] A. Das and S. Bandyopadhyay, "Sentiwordnet for bangla," *Knowledge Sharing Event-4: Task*, vol. 2, 2010.
- [25] A. Das and S. Bandyopadhyay, "Sentiwordnet for indian languages," *Asian Federation for Natural Language Processing, China*, pp. 56–63, 2010.
- [26] D. Das and S. Bandyopadhyay, "Labeling emotion in bengali blog corpus—a fine grained tagging at sentence level," in *Proceedings of the 8th Workshop on Asian Language Resources*, 2010, p. 47.
- [27] D. Narayan, D. Chakrabarti, P. Pande, and P. Bhattacharyya, "An experience in building the indo wordnet—a wordnet for hindi," in *First International Conference on Global WordNet, Mysore, India*, 2002.
- [28] A. Bakliwal, P. Arora, and V. Varma, "Hindi subjective lexicon: A lexical resource for hindi polarity classification," in *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC)*, 2012.
- [29] N. Mittal, B. Agarwal, G. Chouhan, N. Bania, and P. Pareek, "Sentiment analysis of hindi review based on negation and discourse relation," in *Sixth International Joint Conference on Natural Language Processing*, 2013, p. 45.
- [30] V. Jha, N. Manjunath, P. D. Shenoy, and K. Venugopal, "HSRA: Hindi Stopword Removal Algorithm," in *2015 IEEE International WIE Conference on Electrical and Computer Engineering (Wiecon-ECE 2015)*, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, 2015.
- [31] V. Jha, S. R. S. Hebbar, P. D. Shenoy, and V. Kuppanna Rajuk, "HMD-SAD: Hindi Multi-Domain Sentiment Aware Dictionary," in *2015 IEEE International Conference on Computing and Network Communications (CoCoNet'15)*, Trivandrum, Kerala, India, India, 2015.
- [32] H. Yu and V. Hatzivassiloglou, "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*. Association for Computational Linguistics, 2003, pp. 129–136.
- [33] P. Bhattacharyya, "Indowordnet," in *In Proc. of LREC-10*. Citeseer, 2010.
- [34] A. Barrón-Cedeño, A. Eiselt, and P. Rosso, "A comparison of models over wikipedia articles revisions," *ICON*, vol. 2009, 2009.
- [35] V. Jha, N. Manjunath, P. D. Shenoy, K. Venugopal, and L. Patnaik, "Homs: Hindi opinion mining system," in *Recent Trends in Information Systems (ReTIS), 2015 IEEE 2nd International Conference on*. IEEE, 2015, pp. 366–371.