SentiWordNet for Indian Languages

Amitava Das[1] and Sivaji Bandyopadhyay[2]
Department of Computer Science and Engineering
Jadavpur University
amitava.santu@gmail.com[1] sivaji_cse_ju@yahoo.com[2]

Abstract

The discipline where sentiment/ opinion/ emotion has been identified and classified in human written text is well known as sentiment analysis. A typical computational approach to sentiment analysis starts with prior polarity lexicons where entries are tagged with their prior out of context polarity as human beings perceive using their cognitive knowledge. Till date, all research efforts found in sentiment lexicon literature deal mostly with English texts. In this article, we propose multiple computational techniques like, WordNet based, dictionary based, corpus based or generative approaches for generating SentiWordNet(s) for Indian languages. Currently, SentiWordNet(s) are being developed for three Indian languages: Bengali, Hindi and Telugu. An online intuitive game has been developed to create and validate the developed SentiWordNet(s) by involving Internet population. A number of automatic, semi-automatic and manual validations and evaluation methodologies have been adopted to measure the coverage and credibility of the developed SentiWordNet(s).

1 Introduction

Sentiment analysis and classification from electronic text is a hard semantic disambiguation problem. The regulating aspects of semantic orientation of a text are natural language context information (Pang et al., 2002) language properties (Wiebe and Mihalcea, 2006), domain pragmatic knowledge (Aue and Gamon, 2005) and lastly most challenging is the time dimension (Read, 2005).

The following example shows that the polarity tag associated with a sentiment word depends on the time dimension. During 90's mobile phone users generally reported in various online reviews about their color phones but in recent times color phone is not just enough. People are fascinated and influenced by touch screen and various software(s) installation facilities on these new generation gadgets.

In typical computational approaches (Higashinaka et al., 2007; Hatzivassiloglou et al., 2000) to sentiment analysis researchers consider the problem of learning a dictionary that maps semantic representations to verbalizations, where the data comes from opinionated electronic text. Although lexicons in these dictionaries are not explicitly marked up with respect to their contextual semantics, they contain only explicit polarity rating and aspect indicators. Lexicon-based approaches can be broadly classified into two categories firstly where the discriminative polarity tag of lexicons is determined on labeled training data and secondly where the lexicons are manually compiled, the later constitutes the main effective approach.

It is undoubted that the manual compilation is always the best way to create monolingual semantic lexicons, but manual methods are expensive in terms of human resources, it involves a substantial number of human annotators and it takes lot of time as well. In this paper we propose several computational techniques to generate sentiment lexicons in Indian languages automatically and semi-automatically. In the present task, SentiWord-

Proceedings of the 8th Workshop on Asian Language Resources, pages 56–63,
Beijing, China, 21-22 August 2010.

Net(s) are being developed for the Bengali, Hindi and Telugu languages.

Several prior polarity sentiment lexicons are available for English such as SentiWordNet (Esuli et. al., 2006), Subjectivity Word List (Wilson et. al., 2005), WordNet Affect list (Strapparava et al., 2004), Taboada's adjective list (Taboada et al., 2006).

Among these publicly available sentiment lexicon resources we find that SentiWordNet is most widely used (number of citation is higher than other resources[1]) in several applications such as sentiment analysis, opinion mining and emotion analysis. Subjectivity Word List is most trustable as the opinion mining system OpinionFinder[2] that uses the subjectivity word list has reported highest score for opinion/sentiment subjectivity (Wiebe and Riloff, 2006). SentiWordNet is an automatically constructed lexical resource for English that assigns a positivity score and a negativity score to each WordNet synset.

The subjectivity word list is compiled from manually developed resources augmented with entries learned from corpora. The entries in the subjectivity word list have been labeled with part of speech (POS) tags as well as either strong or weak subjective tag depending on the reliability of the subjective nature of the entry.

These two resources have been merged automatically and the merged resource is used for SentiWordNet(s) generation in the present task.

The generated sentiment lexicons or SentiWordNet(s) for several Indian languages mostly contain synsets (approximately 60%) of respective languages. Synset based method is robust for any kind of monolingual lexicon creation and useful to avoid further word sense disambiguation problem in application domain.

Additionally we have developed an online intuitive game to create and validate the developed SentiWordNet(s) by involving Internet population.

The proposed approaches in this paper are easy to adopt for any new language. To measure the coverage and credibility of generated SentiWordNet(s) in Indian languages we have developed several automatic and semi-automatic evaluation methods.

2 Related Works

Various methods have been used in the literature such as WordNet based, dictionary based, corpus based or generative approaches for sentiment lexicon generation in a new target language.

Andreevskaia and Bergler, (2006) present a method for extracting sentiment-bearing adjectives from WordNet using the Sentiment Tag Extraction Program (STEP). They did 58 STEP runs on unique non-intersecting seed lists drawn from manually annotated list of positive and negative adjectives and evaluated the results against other manually annotated lists.

The proposed methods in (Wiebe and Riloff, 2006) automatically generate resources for subjectivity analysis for a new target language from the available resources for English. Two techniques have been proposed for the generation of target language lexicon from English subjectivity lexicon. The first technique uses a bilingual dictionary while the second method is a parallel corpus based approach using existing subjectivity analysis tools for English.

Automatically or manually created lexicons may have limited coverage and do not include most semantically contrasting word pairs like antonyms. Antonyms are broadly categorized (Saif Mohammed, 2008) as *gradable adjectives* (hot–cold, good–bad, friend–enemy) and *productive adjectives* (normal–abnormal, fortune–misfortune, implicit–explicit). The first type contains the semantically contrasting word pairs but the second type includes orthographic suffix/affix as a clue. The second type is highly productive using very less number of affixation rules.

Degree of antonymy (Mohammad et al., 2008) is defined to encompass the complete semantic range as a combined measure of the contrast in meaning conveyed by two antonymy words and is identified by distributional hypothesis. It helps to measure relative sentiment score of a word and its antonym.

Kumaran et al., (2008) introduced a beautiful method for automatic data creation by online intuitive games. A methodology has been

[1] http://citeseerx.ist.psu.edu/
[2] http://www.cs.pitt.edu/mpqa/

proposed for community creation of linguistic data by community collaborative framework known as wikiBABEL[3]. It may be described as a revolutionary approach to automatically create large credible linguistic data by involving Internet population for content creation.

For the present task we prefer to involve all the available methodologies to automatically and semi-automatically create and validate SentiWordNet(s) for three Indian languages. Automatic methods involve only computational methods. Semi-automatic methods involve human interruption to validate system's output.

3 Source Lexicon Acquisition

SentiWordNet and Subjectivity Word List have been identified as the most reliable source lexicons. The first one is widely used and the second one is robust in terms of performance. A merged sentiment lexicon has been developed from both the resources by removing the duplicates. It has been observed that 64% of the single word entries are common in the Subjectivity Word List and SentiWordNet. The new merged sentiment lexicon consists of 14,135 numbers of tokens. Several filtering techniques have been applied to generate the new list.

A subset of 8,427 sentiment words has been extracted from the English SentiWordNet, by selecting those whose orientation strength is above the heuristically identified threshold value of 0.4. The words whose orientation strength is below 0.4 are ambiguous and may lose their subjectivity in the target language after translation. A total of weakly subjective 2652 words are discarded from the Subjectivity word list as proposed in (Wiebe and Riloff, 2006).

In the next stage the words whose POS category in the Subjectivity word list is undefined and tagged as "*anypos*" are considered. These words may generate sense ambiguity issues in the next stages of subjectivity detection. The words are checked in the SentiWordNet list for validation. If a match is found with certain POS category, the word is added to the new merged sentiment lexicon. Otherwise the word is discarded to avoid ambiguities later.

Some words in the Subjectivity word list are inflected e.g., *memories*. These words would be stemmed during the translation process, but some words present no subjectivity property after stemming (*memory* has no subjectivity property). A word may occur in the subjectivity list in many inflected forms. Individual clusters for the words sharing the same root form are created and then checked in the SentiWordNet for validation. If the root word exists in the SentiWordNet then it is assumed that the word remains subjective after stemming and hence is added to the new list. Otherwise the cluster is completely discarded to avoid any further ambiguities.

Various statistics of the English SentiWordNet and Subjectivity Word List are reported in Table 1.

	SentiWordNet		**Subjectivity Word List**	
Entries	**Single**	**Multi**	**Single**	**Multi**
	115424	79091	5866	990
Uambiguous Words	20789	30000	4745	963
Discarded Ambiguous Words	**Threshold**	**Orientation Strength**	**Subjectivity Strength**	**POS**
	86944	30000	2652	928

Table 1: English SentiWordNet and Subjectivity Word List Statistics

4 Target Lexicon Generation

4.1 Bilingual Dictionary Based Approach

A word-level translation process followed by error reduction technique has been adopted for generating the Indian languages SentiWordNet(s) from the English sentiment lexicon merged from the English SentiWordNet and the Subjectivity Word List.

English to Indian languages synsets are being developed under Project English to Indian Languages Machine Translation Systems

[3] http://research.microsoft.com/en-us/projects/wikibabel/

(EILMT)[4], a consortia project funded by Department of Information Technology (DIT), Government of India. These synsets are robust and reliable as these are created by native speakers as well as linguistics experts of the specific languages. For each language we have approximately 9966 synsets along with the English WordNet offset. These bilingual synset dictionaries have been used along with language specific dictionaries.

A word level synset/lexical transfer technique is applied to each English synset/word in the merged sentiment lexicon. Each dictionary search produces a set of Indian languages synsets/words for a particular English synset/word.

4.1.1 Hindi

Two available manually compiled English-Hindi electronic dictionaries have been identified for the present task. First is the SHABDKOSH[5] and the second one is Shabdanjali[6]. These two dictionaries have been merged automatically by replacing the duplicates. The merged English-Hindi dictionary contains approximately 90,872 unique entries. The positive and negative sentiment scores for the Hindi words are copied from their English SentiWordNet.

The bilingual dictionary based translation process has resulted 22,708 Hindi entries.

4.1.2 Bengali

An English-Bengali dictionary (approximately 102119 entries) has been developed using the Samsad Bengali-English dictionary[7]. The positive and negative sentiment scores for the Bengali words are copied from their English SentiWordNet equivalents.

The bilingual dictionary based translation process has resulted in 35,805 Bengali entries. A manual checking is done to identify the reliability of the words generated from automatic process. After manual checking only 1688 words are discarded i.e., the final list consists of 34,117 words.

4.2 Telugu

Charles Philip Brown English-Telugu Dictionary[8], Aksharamala[9] English-Telugu Dictionary and English-Telugu Dictionary[10] developed by Language Technology Research Center (LTRC), International Institute of Hyderabad (IITH) have been chosen for the present task. There is no WordNet publicly available for Telugu and the corpus (Section 4.5) we used is small in size. Dictionary based approach is the main process for Telugu SentiWordNet generation.

These three dictionaries have been merged automatically by replacing the duplicates. The merged English-Telugu dictionary contains approximately 112310 unique entries. The positive and negative sentiment scores for the Telugu words are copied from their English SentiWordNet equivalents.

The dictionary based translation process has resulted in 30,889 Telugu entries, about 88% of final Telugu SentiWordNet synsets. An online intuitive game has been proposed in Section 4.6 to automatically validate the developed Telugu SentiWordNet by involving Internet population.

4.3 WordNet Based Approach

WordNet(s) are available for Hindi[11] (Jha et al., 2001) and Bengali[12] (Robkop et al., 2010) but publicly unavailable for Telugu.

A WordNet based lexicon expansion strategy has been adopted to increase the coverage of the generated SentiWordNet(s) through the dictionary based approach. The present algorithm starts with English SentiWordNet synsets that is expanded using synonymy and antonymy relations in the WordNet. For matching synsets we keep the exact score as in the source synset in the English SentiWordNet. The calculated positivity and negativity score

[4] http://www.cdacmumbai.in/e-ilmt
[5] http://www.shabdkosh.com/
[6] http://www.shabdkosh.com/content/category/downloads/
[7] http://dsal.uchicago.edu/dictionaries/biswas_bengali/
[8] http://dsal.uchicago.edu/dictionaries/brown/
[9] https://groups.google.com/group/aksharamala
[10] http://ltrc.iiit.ac.in/onlineServices/Dictionaries/Dict_Frame.html
[11] http://www.cfilt.iitb.ac.in/wordnet/webhwn/
[12] http://bn.asianwordnet.org/

for any target language antonym synset is calculated as:

$$T_p = 1 - S_p$$

$$T_n = 1 - S_n$$

where S_p, S_n are the positivity and negativity score for the source language (i.e, English) and T_p, T_n are the positivity and negativity score for target languages (i.e., Hindi and Bengali) respectively.

4.3.1 Hindi

Hindi WordNet is a well structured and manually compiled resource and is being updated since last nine years. There is an available API[13] for accessing the Hindi WordNet. Almost 60% of final SentiWordNet synsets in Hindi are generated by this method.

4.3.2 Bengali

The Bengali WordNet is being developed by the Asian WordNet (AWN) community. It only contains 1775 noun synsets as reported in (Robkop et al., 2010). A Web Service[14] has been provided for accessing the Bengali WordNet. There are only a few number of noun synsets in the Bengali WordNet and other important POS category words for sentiment lexicon such as adjective, adverb and verb are absent. Only 5% new lexicon entries have been generated in this process.

4.4 Antonym Generation

Automatically or manually created lexicons have limited coverage and do not include most semantically contrasting word pairs. To overcome the limitation and increase the coverage of the SentiWordNet(s) we present automatic antonymy generation technique followed by corpus validation to check orthographically generated antonym does really exist. Only 16 hand crafted rules have been used as reported in Table 2. About 8% of Bengali, 7% of Hindi and 11% of Telugu SentiWordNet entries are generated in this process.

Affix/Suffix	Word	Antonym
***ab*X**	Normal	*Ab*-normal
***mis*X**	Fortune	*Mis*-fortune
***im*X-exX**	*Im*-plicit	*Ex*-plicit
***anti*X**	Clockwise	*Anti*-clockwise
***non*X**	Aligned	*Non*-aligned
***in*X-*ex*X**	*In*-trovert	*Ex*-trovert
***dis*X**	Interest	*Dis*-interest
***un*X**	Biased	*Un*-biased
***up*X-*down*X**	*Up*-hill	*Down*-hill
***im*X**	Possible	*Im*-possible
***ill*X**	Legal	*Il*-legal
***over*X-*under*X**	Overdone	*Under*-done
***in*X**	Consistent	*In*-consistent
***r*X-*ir*X**	Regular	*Ir*-regular
X*less*-X*ful*	Harm-*less*	Harm-*ful*
***mal*X**	Function	*Mal*-function

Table 2: Rules for Generating Productive Antonyms

4.5 Corpus Based Approach

Language/culture specific words such as those listed below are to be captured in the developed SentiWordNet(s). But sentiment lexicon generation techniques via cross-lingual projection are unable to capture these words. As example:

सहेरा (Sahera: A marriage-wear)

দুর্গাপুজো (Durgapujo: A festival of Bengal)

To increase the coverage of the developed SentiWordNet(s) and to capture the language/culture specific words an automatic corpus based approach has been proposed. At this stage the developed SentiWordNet(s) for the three Indian languages have been used as a seed list. Language specific corpus is automatically tagged with these seed words and we have a simple tagset as SWP (Sentiment Word Positive) and SWN (Sentiment Word Negative). Although we have both positivity and negativity scores for the words in the seed list but we prefer a word level tag as either positive or negative following the highest sentiment score.

A Conditional Random Field (CRF[15]) based Machine Learning model is then trained with the seed list corpus along with multiple linguistics features such as morpheme, parts-of-

[13] http://www.cfilt.iitb.ac.in/wordnet/webhwn/API_downloaderInfo.php

[14] http://bn.asianwordnet.org/services

[15] http://crfpp.sourceforge.net

speech, and chunk label. These linguistics features have been extracted by the shallow parsers[16] for Indian languages. An n-gram (n=4) sequence labeling model has been used for the present task.

The monolingual corpuses used have been developed under Project English to Indian Languages Machine Translation Systems (EILMT). Each corpus has approximately 10K of sentences.

4.6 Gaming Methodology

There are several motivations behind developing an intuitive game to automatically create multilingual SentiWordNet(s). The assigned polarity scores to each synset may vary in time dimension. Language specific polarity scores may vary and it should be authenticated by numbers of language specific annotators.

In the history of Information Retrieval research there is a milestone when ESP[17] game (Ahn et al., 2004) innovate the concept of a game to automatically label images available in World Wide Web. Highly motivated by the historical research we proposed a intuitive game to create and validate SentiWordNet(s) for Indian languages by involving internet population.

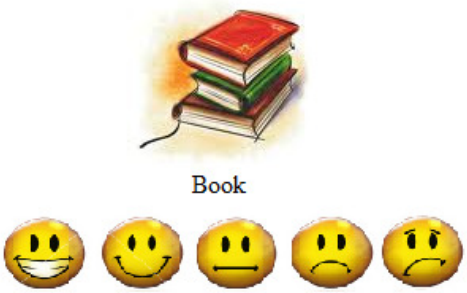

Figure 1: Intuitive Game for SentiWordNet(s) Creation

In the gaming interface a simple picture (retrieved by Google Image API[18]) along with a sentiment bearing word (retrieved randomly from SentiWordNet) is displayed to a player and he/she is then been asked to capture his immediate sentiment as extreme positive, positive, extreme negative, negative or neutral by pressing appropriate emoticon buttons. A snap of the game is shown in the Figure 1. The sentiment score is calculated by the different emoticons based on the inputs from the different players and then is assigned the scale as follows: extreme positive (pos: 0.5, neg: 0.0), positive (pos: 0.25, neg: 0.0), neutral (pos: 0.0, neg: 0.0), negative (pos: 0.0, 0.25), extreme negative (pos: 0.0, neg: 0.5).

The score of a particular player is calculated on the basis of pre-stored sentiment lexicon scores in the generated SentiWordNet(s).

5 Evaluation

Andera Esuli and Fabrizio Sebastiani (2006) have calculated the reliability of the sentiment scores attached to every synsets in the English SentiWordNet. They have tagged sentiment words in the English WordNet with positive and negative sentiment scores. In the present task, these sentiment scores from English WordNet have been directly copied to the Indian language SentiWordNet(s).

Two extrinsic evaluation strategies have been adopted for the developed Bengali SentiWordNet based on the two main usages of the sentiment lexicon as subjectivity classifier and polarity identifier. The Hindi and Telugu SentiWordNet(s) have not completely been evaluated.

5.1 Coverage

	NEWS	BLOG
Total number of documents	100	-
Total number of sentences	2234	300
Avgerage number of sentences in a document	22	-
Total number of wordforms	28807	4675
Avgerage number of wordforms in a document	288	-
Total number of distinct wordforms	17176	1235

Table 3: Bengali Corpus Statistics

We experimented with NEWS and BLOG corpora for subjectivity detection. Sentiment lexicons are generally domain independent but it provides a good baseline while working with sentiment analysis systems. The coverage of

[16] http://ltrc.iiit.ac.in/showfile.php?filename=downloads/shallow_parser.php

[17] http://www.espgame.org/

[18] http://code.google.com/apis/ajaxsearch/multimedia.html

the developed Bengali SentiWordNet is evaluated by using it in a subjectivity classifier (Das and Bandyopadhyay, 2009). The statistics of the NEWS and BLOG corpora is reported in Table 3.

For comparison with the coverage of English SentiWordNet the same subjectivity classifier (Das and Bandyopadhyay, 2009) has been applied on Multi Perspective Question Answering (MPQA) (NEWS) and IMDB Movie review corpus along with English SentiWordNet. The result of the subjectivity classifier on both the corpus proves that the coverage of the Bengali SentiWordNet is reasonably good. The subjectivity word list used in the subjectivity classifier is developed from the IMDB corpus and hence the experiments on the IMDB corpus have yielded high precision and recall scores. The developed Bengali SentiWordNet is domain independent and still its coverage is very good as shown in Table 4.

Languages	Domain	Precision	Recall
English	**MPQA**	76.08%	83.33%
	IMDB	79.90%	86.55%
Bengali	**NEWS**	72.16%	76.00%
	BLOG	74.6%	80.4%

Table 4: Subjectivity Classifier using SentiWordNet

5.2 Polarity Scores

This evaluation metric measures the reliability of the associated polarity scores in the sentiment lexicons. To measure the reliability of polarity scores in the developed Bengali SentiWordNet, a polarity classifier (Das and Bandyopadhyay, 2010) has been developed using the Bengali SentiWordNet along with some other linguistic features.

Features	Overall Performance Incremented By
SentiWordNet	**47.60%**

Table 5: Polarity Performance Using Bengali SentiWordNet

Feature ablation method proves that the associated polarity scores in the developed Bengali SentiWordNet are reliable. Table 5 shows the performance of a polarity classifier using the Bengali SentiWordNet. The polarity wise overall performance of the polarity classifier is reported in Table 6.

Polarity	Precision	Recall
Positive	56.59%	52.89%
Negative	75.57%	65.87%

Table 6: Polarity-wise Performance Using Bengali SentiWordNet

Comparative study with a polarity classifier that works with only prior polarity lexicon is necessary but no such works have been identified in literature.

An arbitrary 100 words have been chosen from the Hindi SentiWordNet for human evaluation. Two persons are asked to manually check it and the result is reported in Table 7. The coverage of the Hindi SentiWordNet has not been evaluated, as no manually annotated sentiment corpus is available.

Polarity	Positive	Negative
Percentage	88.0%	91.0%

Table 7: Evaluation of Polarity Score of Developed Hindi SentiWordNet

For Telugu we created a version of the game with Telugu words on screen. Only 3 users have played the Telugu language specific game till date. Total 92 arbitrary words have been tagged and the accuracy of the polarity scores is reported in Table 8. The coverage of Telugu SentiWordNet has not been evaluated, as no manually annotated sentiment corpus is available.

Polarity	Positive	Negative
Percentage	82.0%	78.0%

Table 8: Evaluation of Polarity Score of Developed Telugu SentiWordNet

6 Conclusion

SentiWordNet(s) for Indian languages are being developed using various approaches. The game based technique may be directed towards a new way for the creation of linguistic data not just only for SentiWordNet(s) but in either areas of NLP too.

Presently only the Bengali SentiWordNet[19] is downloadable from the author's web page.

[19] http://www.amitavadas.com/sentiwordnet.php

References

Andreevskaia Alina and Bergler Sabine. CLaC and CLaC-NB: Knowledge-based and corpus-based approaches to sentiment tagging. In Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), pages 117–120, Prague, June 2007.

Aue A. and Gamon M., Customizing sentiment classifiers to new domains: A case study. In Proceedings of Recent Advances in Natural Language Processing (RANLP), 2005.

Das A. and Bandyopadhyay S. (2010). Phrase-level Polarity Identification for Bengali, In International Journal of Computational Linguistics and Applications (IJCLA), Vol. 1, No. 1-2, Jan-Dec 2010, ISSN 0976-0962, Pages 169-182.

Das A. and Bandyopadhyay S. Subjectivity Detection in English and Bengali: A CRF-based Approach. In the Proceeding of ICON 2009.

Esuli Andrea and Sebastiani Fabrizio. SentiWordNet: A publicly available lexical resource for opinion mining. In Proceedings of Language Resources and Evaluation (LREC), 2006.

Hatzivassiloglou, Vasileios and Wiebe Janyce. Effects of adjective orientation and gradability on sentence subjectivity. In Proceedings of COLING-00, 18th International Conference on Computational Linguistics. Saarbru¨cken, GE. Pages 299-305. 2000.

Higashinaka Ryuichiro, Walker Marilyn, and Prasad Rashmi. Learning to generate naturalistic utterances using reviews in spoken dialogue systems. ACM Transactions on Speech and Language Processing (TSLP), 2007.

Jha S., Narayan D., Pande P. and Bhattacharyya P. A WordNet for Hindi, International Workshop on Lexical Resources in Natural Language Processing, Hyderabad, India, January 2001.

Kumaran A., Saravanan K. and Maurice Sandor. WikiBABEL: Community Creation of Multilingual Data, in the WikiSYM 2008 Conference, Porto, Portugal, Association for Computing Machinery, Inc., September 2008.

Mihalcea Rada, Banea Carmen and Wiebe Janyce. Learning multilingual subjective language via-cross-lingual projections. In Proceedings of the Association for Computational Linguistics (ACL), pages 976–983, Prague, Czech Republic, June 2007.

Mohammad Saif, Dorr Bonnie, and Hirst Graeme. Computing Word-Pair Antonymy. In Proceedings of the Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-2008), October 2008, Waikiki, Hawaii.

Pang Bo, Lee Lillian, and Vaithyanathan Shivakumar. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79–86, 2002.

Read Jonathon. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In Proceedings of the ACL Student Research Workshop, 2005.

Robkop Kergrit, Thoongsup Sareewan, Charoenporn Thatsanee, Sornlertlamvanich Virach and Isahara Hitoshi.WNMS: Connecting the Distributed WordNet in the Case of Asian WordNet. . In the Proceeding of 5th International Conference of the Global WordNet Association (GWC-2010), Mumbai, India , 31st Jan. - 4th Feb., 2010.

Wiebe Janyce and Mihalcea Rada. Word sense and subjectivity. In Proceedings of COLING/ACL-06 the 21st Conference on Computational Linguistics/Association for Computational Linguistics. Sydney, Australia. Pages 1065--1072.

Wiebe Janyce and Riloff Ellen. Creating Subjective and Objective Sentence Classifiers from Unannotated Texts. In Proceeding of International Conference on Intelligent Text Processing and Computational Linguistics, Mexico City, Pages 475–486, 2006.

Wilson Theresa, Wiebe Janyce and Hoffmann Paul (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In Proceedings of HLT/EMNLP 2005, Vancouver, Canada.