Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 93 (2016) 554 – 563

Procedia Computer Science

6th International Conference On Advances In Computing & Communications, ICACC 2016, 6-8 September 2016, Cochin, India

Context Specific Lexicon for Hindi Reviews

Deepali Mishra[a], Manju Venugopalan[a] and Deepa Gupta[b,*]

[a]*Department of Computer Science, Amrita School of Engineering, Bangalore, Amrita Vishwa Vidyapeetham, Amrita University, India, deepalitiwari22@gmail.com*
[a]*Department of Computer Science , Amrita School of Engineering, Bangalore, Amrita Vishwa Vidyapeetham, Amrita University, India, v_manju@blr.amrita.edu*
[b]*Department of Mathematics, Amrita School of Engineering, Bangalore, Amrita Vishwa Vidyapeetham, Amrita University, India, g_deepa@blr.amrita.edu*

Abstract

In the era of social networking, immense amount of posts, comments and tweets generated every second are increasing the size of social database .The analysis of this voluminous data is necessary for exploring the orientation of people's opinion about a particular entity. Most of the online data are in English language, but due to increase in technology and improved awareness of people, the online data available in Indian languages are gradually increasing. Sentiment analysis of English language alone is not sufficient to know the inclination of people towards an entity, other Indian language sentiment analysis is a must, their contribution is also important for us. The available sentiment classification lexicon resources like Hindi SentiWordNet are generic in nature and hence results in average sentiment classification accuracy due to contextual dependency. To improve the sentiment classification accuracy, we present an improvised lexicon resource for Hindi language for Hotel and Movie domains. The improvised polarity lexicon has been built reflecting context sensitivity and to increase coverage it has been expanded used synonyms based approach. The built polarity lexicon resource showcases an improvement in accuracy of 42% and 78% in Movie and Hotel domain, respectively, compared to the existing Hindi SentiWordNet lexicon resource.

Peer-review under responsibility of the Organizing Committee of ICACC 2016

Keywords: Sentiment analysis ; lexicon ; HSWN ; LR ; LRE etc .

* Deepa Gupta. Tel.:+919916921850.
E-mail address: g_deepa@blr.amrita.edu

Peer-review under responsibility of the Organizing Committee of ICACC 2016
doi:10.1016/j.procs.2016.07.283

1. Introduction

The current decade has been witnessing an exponential increase in the number of users and web content. This voluminous data are used by people to get an idea in decision making about any entity. For example before travelling to any unknown place, previously we would prefer talking to those who have visited that place, but now due to online available data in the form of reviews, we go by the reviews for a decision making. These available text data need to be analyzed, and hence the opinion orientation identified which is termed as opinion mining or sentiment classification. Almost two decades of work has been contributed to extracting sentiment from English the broader categories being sentiment classification, lexicon resource creation etc. but minimal work have happened on Indian languages. The increase in the volume of Indian language data available online has elevated the importance of exploring sentiment in Indian Languages.

With the advent of technology where many social networking sites like Twitter, Facebook etc. providing provisions to express in a handful of Indian Languages, newspapers, blogs etc. providing provisions for native expressions have led to more Indian Language content available online. Even though English is an International Language, the sentiment extracted from English reviews alone cannot be considered to make final conclusions on an entity; other language inputs should also be considered. This creates the necessity to give some effort to sentiment analysis of Regional Languages.

The last few years have witnessed some authors showing their interest to mining in Indian languages but as mentioned earlier majority work contributions are in English. So it is obvious that more resources and tools are available for the same. Hindi is a well-known and widely spoken language in India. Web pages in Hindi language have increased on a rapid pace. There are many websites which provide information in Hindi owned by various news websites providing information regarding culture, music, entertainment and other aspects of arts. The web content for Hindi language has been increasing with great speed. This emphasizes the scope for further exploration of the language. But each language puts forward challenges to be encountered in terms of its syntactic and semantic structures. Hindi is a free order language with various morphological variants, spelling variance, word sense ambiguity and contextual variances. Sentiment analysis in Hindi is less explored so there is scarcity of resources and tools. Among the existing resources the most popularly used is the Hindi SentiwordNet[1]. The classification based research works using this resource have found to exhibit average accuracy which owes to the polarity lexicons not being context sensitive. Opinion words might infer different meanings in varied domains. For example "**इस सैमसंग मोबाइल की बैटरी लाइफ लंबी हैं |**", "**फिल्म लंबी थी |**". In the first sentence the "**लंबी**" word in battery life context expresses a positive opinion, but in the second sentence "**लंबी**" word in movie context conveys a negative opinion. The polarity of the word contributed by Hindi SentiwordNet is +0.5 which is sensible for the cellphone battery context but not for the movie domain. Hence this work takes a special interest towards dealing with context specificity issue. The major contributions put forward by the proposed work are

a) Proposes an algorithm to build an improvised context sensitive polarity lexicon for a particular domain.
b) Attempts improving the lexicon coverage by the Hindi WordNet based approach

The research works attempted in Hindi Sentiment analysis have been keenly studied and the findings presented in Section 2.The Corpus details are provided in Section 3, the detailed Proposed Approach in Section 4, the Results and Analysis in Section 5 and the Conclusion and Future work in Section 6.

2. Related Works

The earliest works in Hindi Sentiment analysis can be traced back to the beginning of the current decade. Most of the works attempted classification on different domains using existing resources like Hindi SentiWordNet[1]. The work has contributed SentiWordNet for the 3 Indian languages Hindi, Bengali and Telugu by using the English SentiWordNet and the subjective word list as base resource. To build the lexicon resources for target language, the experimented approaches are machine translation or dictionary based, word net based, corpus based and online game based[2]. English SentiWordNet words are translated into target language and the same polarity score has been given to target language lexicons. To increase the lexicons in the generated target language SentiWordNet used the wordNet based approach in which synonym of a word has given same polarity score and antonym has given opposite polarity score. All the built lexicon resources have been evaluated manually. Classification accuracy enhancement has been attempted by authors using different algorithms like negation handling, word replacement and machine translation methods. In negation handling, the opposite polarity is assigned to opinion words for the presence of negation words like न, नहीं etc. within a predefined window size. Word replacement algorithm assigns a word, the polarity score of its synonym if the word is not present in the Hindi SentiWordNet. In the Machine translation method, if a word is not in the Hindi SentiWordNet, its English equivalent score is fetched from English SentiWordNet and by this score sentiment classification is attempted. These different methodologies provide accuracy in the average range, the reason being the most widely used resource Hindi SentiWordNet is built by machine translation approach, so the lexicon polarity is not context sensitive and the broader challenges put forward by Hindi language in word sense ambiguity. To refine the polarity score of lexicons, a work has been attempted in [3] using a graph based WorldNet approach but the built lexicon resource contains only adjective and adverbs. They claim that their lexicon resource renders 79% accuracy but it doesn't address contextual sensitivity. In [4] efforts has been made to find the correct polarity score of lexicon according to the context .They built a vector space model by using the semantic net and SentiWordNet for Bengali language . They used the Bengali news corpus and have reported 70% accuracy by presented approach.

As compared to lexicon resource generation more work has been explored in sentiment classification. For the classification work HSWN (Hindi SentiWordNet) is used. In [5] sentiment classification result has been increased by using an improved HSWN, negation handling and discourse. Improved HSWN is made by using machine translation method i.e. if a word is not in HSWN than the word is translated into English and the translated word polarity score from the SWN is coined to the original Hindi token. In negation handling they have targeted the negative words in Hindi which appear before and after a word or combination of word and hence change the meaning of sentence. To attack this situation they had described solution as assigning the opposite polarity to lexicon word preceded by a negative word. Discourses are those words like मगर, लेकिन, बावजूद etc. which gives more weightage to specific parts of the sentence. The work identifies the discourse and according to the word inclination in the sentence they have done the sentiment classification. The combined techniques fetched them 80.21 % classification accuracy in the movie review on test data. In [6] they used the word replacement approach to increase the classification accuracy. If a word is not in the found in HSWN than the word is replaced by the same meaning word that is present in the HSWN and hence a polarity score which contributes to sentiment classification. The authors [7] have performed sentiment classification and text normalization on the review and feedback data collected from Facebook and YouTube. The data contained text written in both Language Hindi and English. They used lexicon based approach for the SA and trained the classifier for handling abbreviations, Wordplay, Slang word and phonetic typing. They have performed language identification on sentences and translated Hindi words written in English to Hindi Devnagari script. For Sentiment classification of English, the Opinion Lexicon and AFINN list has been used, Hindi SentiWordNet for Hindi data. Sentiment Classification performed on positive, negative and neutral categories and neutral reviews are reclassified by using WorldNet based approach and the work claimed accuracy above 85%. In [8], sentiment classification experimented by three approaches In-language, Machine translation and Resource based approach. They manually annotated the Hindi movie corpus for this work. They have reported an accuracy of 78.14 using the In-language sentiment analysis. In [9] they have explored the Sentiment analysis work in one more direction called Cross-Lingual Sentiment Analysis, here one language test data sentiment analysis done by the lexicon resource build in other language and this of work mostly done by Machine translation method, but here they proposed a supervised sentiment classification approach using word sense as feature .the work has been done for Hindi and Marathi language. In this approach, first they found the words from two languages from both language WorldNet which are used for one concept in both language and included the synonyms of the word and gave the same synset identifier to both language words for one concept, by this way they created a common corpus as lexicon resource and done cross-lingual sentiment classification. They adapted travel destination reviews for classification work and claimed accuracy of 72% and 84% for Hindi and Marathi sentiment classification respectively. [10] performed the Real time sentiment analysis in tweets data by using supervised approach and the tweets are about the AAP party and Python language. They build a polarity lexicon using Stanford university tweet data set. They build two naïve based classifier with some variation like baseline classifier is trained with original tweet data with label positive, negative and neutral and second is trained with positive and negative data. Sentiment classification experimented with different features and got average accuracy. [11] has proposed a model for sentiment classification on Hindi tweets. Multinomial Naïve Bayes method has been applied for classification and showcases an average accuracy of 50.75% .The proposers of [12] have contributed a benchmark dataset for the Aspect level sentiment analysis for Hindi language. They have collected data from 12 domains sourced from different websites and manually annotated the reviews, in which they have annotated the aspect term, aspect term category, aspect term polarity and classified the sentences into categories positive, negative ,neutral and conflict. They used the conditional random field model using different features like Word & local context, POS information, chunk information, suffix and prefix information for the aspect term extraction and the SVM model for the sentiment analysis. A survey on the various works carried out in Sentiment Analysis of Hindi language [13] categorizes them into two broad areas, lexicon resource creation and sentiment classification. The approaches, techniques, limitations and accuracy attained in the various explored methods have been presented.

The work in [14] has been dedicated towards phrase level polarity detection in Bengali language. For this news data set has been used and classified as subjective data by using subjective classifier. They used hybrid approach for phrase level polarity detection. They extracted the phrase adapting the lexicon entities and linguistic syntactic features and evaluated the result which shows a precision of 70.04% and recall of 63.02%. [15] aims to resolve context sensitive issues by building domain specific and domain independent lexicon resources. Datasets were chosen from different domains which are product reviews by customers. The idea was to incorporate the contextual learning knowledge on multiple domains in the form of domain independent and domain specific lexicons. The approach contributed to significant improvement of around 8 points beyond the SentiWordNet baseline. The proposed work has drawn insights from [15].

Most of the existing lexicon creation approaches are translation based and hence had to compromise in the result obtained. The coverage of these lexicons is hardly contributing to 60%. Minimal works have incorporated contextual polarity. This highlights the importance of polarity lexicons which are context sensitive. Hence this work is focused on building an effective context sensitive polarity lexicon for a particular domain.

3. Proposed Approach

The work aims to build a domain specific dictionary for the chosen domain. The phases involved in lexicon generation are presented as different modules the Opinion word extraction module, the Context Specific Polarity Lexicon (CSPL) Building module and the CSPL extension module. The phases involved in lexicon generation are depicted in Fig. 1.

3.1 Opinion Word Extraction Module

The input raw data in the form of customer reviews are fed through a pre-processing stage. In the pre-processing stage, the collected review data is cleaned which involves the removal of punctuation like symbols, spell check and tokenization(which refers to splitting the review into sentences and sentences further into words), POS tagging(assigns Part of Speech tag like NN for noun, JJ for Adjective) and lemmatization(reducing to root word). For tokenization and POS tagging, the Hindi POS Tagger 3.0(http://sivareddy.in/downloads) has been used. A sample output of used POS tagger is displayed in Fig. 2.

Each review output is presented in a predefined format of the used POS Tagger. In each output line, the first word represents the original word in the review, second word shows the root word of the original word and the third word gives the POS tag of original word. The fifth word refers to the broad class of the POS tag. The remaining part of the output line does not contribute to the proposed work. The output is characterized by different POS tags like QF as Quantifier, NN as Noun, JJ as adjective, NEG as negative and VM as verb.

The pre-processing part outputs all root words in the reviews tagged by their corresponding POS tags. The words with POS tags under the broad classes of Nouns(except proper nouns tagged NNP), Verbs(except auxiliary verbs tagged VAUX), Adverbs and Adjectives alone are considered as opinion oriented words in the proposed work.

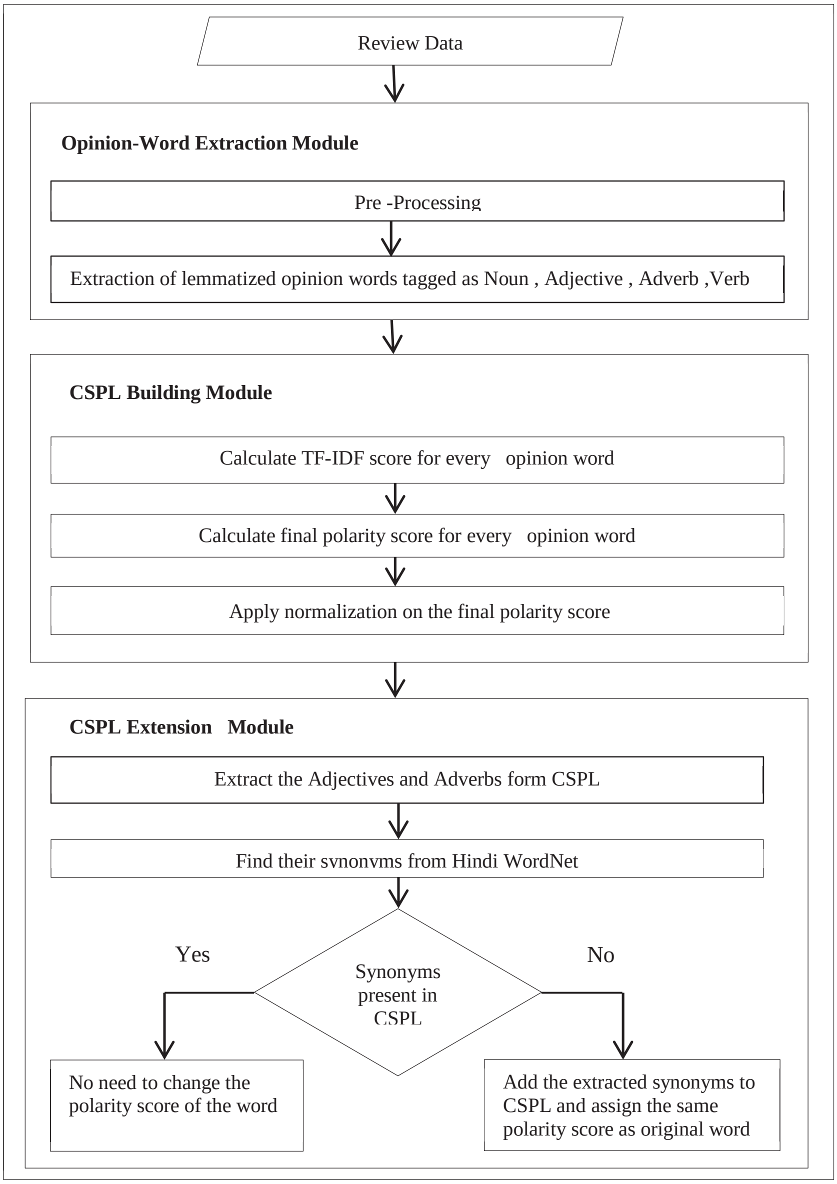

Fig. 1 Schematic diagram of Context Specific Polarity Lexicon (CSPLE) Building

कुछ	कुछ	QF	-	adj	any	any	-	d
खास	खास	JJ	-	adj	any	any	-	any
फिल्म	फिल्म	NN	0	n	f	sg	3	d
नहीं	नहीं	NE	-	adv	-	-	-	-
हैं	है	VM	है	v	any	pl	1G	-

Fig. 2 Sample output of used POS tagger

3.2 Context Specific Polarity Lexicon (CSPL) Building Module

The opinion words extracted in the previous module are assigned a polarity score in this module. The popular method TF-IDF which is a statistical measure of inclination of every token towards any one of the classes is the indexing method used. The frequency of opinion words in both classes of reviews i.e. in positive and negative reviews, the number of reviews in which a particular lexicon is found all these serves as contributor to the final score. Formula (1) is used to calculate the TF-IDF of each opinion word.

$$fp\ (w)\ =\ loge\ (freq\ (w)\ +1)\ *\ loge\ (\ N\ /\ rf\ (w)\ ) \quad (1)$$

In the above formula the term *fp(w)* refers to the TF-IDF score, *freq(w)* expresses the number of times a token *w* occurred in individual reviews and the term *rf(w*) shows the count of reviews in which lexicon *w* is seen . *N* shows the total number of reviews taken for building the Polarity Lexicon. The final polarity score of each opinion word *dfp(w)* is calculated by shown formula (2).

$$dfp\ (w) = (\ fp(w) in\ positive\ reviews\ ) -\ (\ fp(w)\ in\ negative\ reviews\ ) \quad (2)$$

The final polarity scores of the opinion words are subjected to normalization as the proposed work attempts variations where the built CSPL is supported by Hindi SentiWordnet and hence would require both the set of values to confine to the same range. The normalisation is performed separately for each POS tag. HSWN polarity score vary between -1 and +1. Each word is normalized by its maximum value score of POS tag with the polarity score sign. Normalization is bounded according to the POS tag implemented with the aim that polarity score is biased with their POS tag only. For e.g. if word comes under the category of adverb and word score has a negative polarity than it is normalized by the maximum value of adverb word score with negative sign. This method had been adopted to build a Normalized Polarity score corpus which forms the Context Specific Polarity Lexicon(CSPL).

3.3 *Context Specific Polarity Lexicon Extension (CSPLE) Module*

To increase the coverage of built Lexicon Resource, Hindi WordNet based approach is used. The opinion words tagged as Adverbs and Adjectives alone in CSPL are extracted. All the synonyms of these extracted words are found from Hindi WordNet. If any synonym of word with the same POS tag value already exists in the corpus then the polarity score of that word is unaltered else the same polarity score is assigned to the synonyms and it is added to the lexicon. In a scenario where a word and one or more of its synonym already exists in the CSPL, a new word which is extracted as a synonym will be assigned a value which is maximum among the existing words in CSPL with the same meaning. Applying this method we increase the coverage of CSPL and the extended resource is referred to as Context Specific Polarity Lexicon with Synonym Extension (CSPLE).

4. Corpus Details and Experimental setup

The dataset has been built by collecting Hindi Movie`s reviews from NavBharatTimes Online news journal and the Hotel reviews from goibibo online travel website. The reviews for the hotel domain were originally in English and have been translated to Hindi using Google translator for our work. The translated reviews are subjected to a post editing phase for rectifying incorrect structural formats and has been done manually. The reviews are labelled data, the review rating expressed between 1 and 5. Here we have segregated reviews rated in the range 3.5 to 5 as positive and 1 to 2.5 rated reviews as negative. Our data consists of 5200 reviews from both the Movie and Hotel domain in which 5000 reviews (2500 +ve and 2500 –ve) are for creating the lexicon resource and the rest 200 for testing the Built Context Specific Polarity Lexicon(CSPL) resource. The corpus statistics are presented in Table 1.

Table 1. Corpus statistics

Domain	Maximum no of sentences in a review	Minimum no of sentences in a review	Average review length	Total no. of sentences
Movie	6	1	2	7950
Hotel	20	2	5	10359

The built Context specific polarity Lexicon (CSPL) resource improvement has been experimented in four variations and has been compared to the Hindi SentiWordNet baseline. The four variations include :

a) Hindi SentiWordNet (HSWN)
In this model, we used the polarity scores given by Hindi SentiWordNet for Sentiment classification of test data. According to the root word POS tag the polarity score is fetched from Hindi SentiWordNet (HSWN).

b) Context Specific Polarity Lexicon(CSPL)
In this model, we used the polarity scores of the built Context specific polarity lexicon without Synonym extension (CSPL).

c) Context Specific Polarity Lexicon and Hindi SentiWordNet(CSPL+HSWN)
In this model, first used the Context specific polarity lexicon without Synonym extension (CSPL) for fetching the polarity score and the unfound lexicon scores are fetched from Hindi SentiWordNet (HSWN).

d) Context Specific Polarity Lexicon with Polarity Extension (CSPLE)
In this Model, the built Context specific polarity Lexicon with synonym Extension (CSPLE) is the only source for obtaining polarity Scores of lexicons.

e) Context Specific Polarity Lexicon with Synonym Extension and Hindi SentiWordNet (CSPLE + HSWN)
In this model, first used the Context specific polarity lexicon with Synonym extension (CSPLE) for getting the polarity score and the unfound lexicon scores are fetched from Hindi SentiWordNet (HSWN).

The models are tested on unknown 200 reviews each on both the domains to test the efficiency of the polarity lexicon created. Hindi SentiWordNet has been used as the baseline for performance evaluation. The performance of CSPL and its variations are measured using the metrics Accuracy, Specificity(proportion of correctly classified positive instances) and Sensitivity(proportion of correctly classified negative instances).

5. Results and Analysis

The results of implementing the proposed approach have been presented in this Section. As observed from Table 1, the average review length in terms of sentences in movie reviews is small compared to Hotel domain owing to the source being original and translated data respectively. Reviews in local languages are found to be less expressive when compared to English which also contributes to the observation. Table 2 showcases the number of opinion words in CSPL and CSPLE under each POS tag . The number of opinion words in the Hotel domain is less compared to that of the movie domain. This might be attributed to the fact that the variety of words in pure language reviews would be more than translated reviews which usually are framed by commonly used words. By the synonym extension approach the increase in the coverage is more in the Hotel domain as compared to Movie domain.

The results of testing across all the models have been displayed in Table 3. The accuracy of classification has been the best in Movie domain for CSPLE and in the Hotel domain CSPL outperformed other models. The performance of the proposed approach has also been measured in terms of its Specificity and Sensitivity depicted in Table 3. CSPL and its variations have been observed to be more specific than sensitive. The built CSPL has shown an improvement of around 42% in Movie domain and 78% in the Hotel domain respectively. Synonyms Extension (SE) and being supported by HSWN which are methods to improve the coverage, yielded positive results in the Movie domain but the hotel domain showed a dip in the accuracy score. SE brought about 5% increase in accuracy in the movie domain. The unexpected result in the hotel domain on applying SE could be attributed to the fact that the Hotel reviews have been derived from translated English source. The corpus created from translated data are usually characterized by commonly used words and a synonym extension approach in this scenario adds pure and varied language words which need not contribute to improving sentiment classification accuracy.

Table 2 .Details of Context Specific Polarity lexicon in Movie and Hotel domain

Model name	Noun		Adjective		Verb		Adverb		Total	
	Movie	Hotel	Movie	Hotel	Movie	Hotel	Movie	Hotel	Movie	Hotel
CSPL	2631	1969	1049	960	532	322	38	34	4251	3285
CSPLE	2631	|1969	9169	11911	532	322	564	603	12896	14805

Table 3.Result of Sentiment classification across models

Model name	Accuracy (%)		Sensitivity		Specificity	
	Movie	Hotel	Movie	Hotel	Movie	Hotel
HSWN	52.5	46.0	0.27	0.81	0.76	0.72
CSPL	71.0	88.0	0.81	0.86	0.66	0.90
CSPL + HSWN	76.5	85.0	0.81	0.79	0.72	0.91
CSPLE	77.0	82.5	0.85	0.82	0.69	0.83
CSPLE + HSWN	75.0	81.5	0.79	0.80	0.71	0.83

To make a comparison among the different models in terms of their coverage capabilities, Table 4 has been presented, which shows each model coverage on the test data.

Table 4.Coverage statistics, comparison between HSWN , CSPL and CSPLE Models

Domain	No. of corpus words	Words covered by HSWN	Words covered by CSPL	Words covered by CSPLE
Movie	833	108	477	703
Hotel	1356	330	782	963

From the Testing corpus, a few reviews from both the domains are presented in Table 5 and thereby compare its polarity score as derived from HSWN and CSPLE.

Table 5. Comparison of HSWN Model and CSPLE Model by polarity scores attained

Domain	Review	Class label	Total score by HSWN	Total score by CSPLE
Movie	फिल्म बहुत अच्छी है. अजय देवगन और सभी कलाकारो ने बहुत अच्छा अभिनय किया है मार धाड की फिल्मो से अलग , फिल्म अपना प्रभाव छोड़ती है. पूरे परिवार के साथ आप फिल्म का आनंद ले सकते है।	Positive	-1.25	0.933
Movie	अच्छी फिल्म , निर्देशन और अभिनय सराहनीय है। लेकिन बीच में कहानी कमजोर और उबाऊ है। अतिरिक्त भावनाएँ , दर्शकों की रुचि कम करती हैं , एक बार देख सकते हैं।	Negative	-0.125	-0.316
Hotel	होटल रेलवे स्टेशन के पास स्थित है और हवाई अड्डे से मुफ्त पिक करता है। कमरे आधुनिक और बहुत विशाल थे। स्टाफ विनम्र और मिलनसार था।	Positive	-0.75	1.08
Hotel	कमरे की चाबी लेने के लिए काफी लंबा इंतजार किया।	Negative	0.125	-0.324

The fourth example from Table 5 from Hotel domain "कमरे की चाबी लेने के लिए काफी लंबा इंतजार किया |". In the given example if HSWN is used for classification then according to the most commonly used polarity score of HSWN the polarity score returned for words काफी and लंबा are 0.25 and -0.125 and hence the review is classified as a positive which is incorrect for the context movie. But CSPL model covers the words कमरे, चाबी, लंबा and इंतजार with polarity score values 0.3 , -0.399, -0.04 and -0.191 respectively, the total score being -0.324 and hence the review classified as negative. The better performance of CSPLE and its effective coverage is clear from the example. Multiple synsets are returned by HSWN for the polarity scores of the word लंबा under the same POS tag. Any of these results do not classify the review as negative.

The improved results obtained experimenting CSPL and its variations have proved the context specificity of the model. A pinpointing thing that has been observed in HSWN is that around 2096 synset ids have polarity score [0.0 0.0] and its synonyms too would be treated neutral. This would mean that 69% of the synset ids convey a neutal sentiment and this accounts for the minimal coverage of HSWN for any domain.

The improvement in accuracy of the proposed model could have been constrained by the fact that the performance of this model is inturn dependent on the quality of the POS tagger and spelling variations in Hindi language.

6. Conclusion& Future work

By adapting the Context Specific Polarity Lexicons the Sentiment classification accuracy is 77% in the Movie Domain for CSPLE model and 88% in Hotel Domain for CSPL model. The source of data being machine translated from English in the Hotel Domain could have been a reason for the accuracy to be effected. Translated data often causes loss of contextuality . In this work, we have increased the coverage of the Polarity Lexicon by including the synonyms of Adjectives and Adverbs. The unigram model has been used. The poor performance of synonym extension in Hotel domain could be due to those synonyms that are not contextually appropriate according to the score.

The future work should focus towards the improvement of the Context Specific Polarity Lexicon and hence the classification accuracy. Larger datasets and antonyms extensions are improvisations to be applied to Polarity Lexicon. Negation handling and experimenting with bigram and trigram models are enhancements in the classification procedure.

References

1. A.Das, S. Bandyopadhyay. SentiWordnet for Indian Languages. In: Proceedings of the 8th Workshop on Asian Language Resources, 2010, pp. 56–63
2. A.Das, S. Bandyopadhay. Dr Sentiment Creates SentiWordNet(s) for Indian Languages Involving Internet Population. In: Proceedings of Indo- wordnet workshop, 2010.
3. Akshat Bakliwal, Piyush Arora , Vasudeva Varma. Hindi subjective lexicon: A lexical resource for Hindi polarity classification. In: Proceedings of the Eighth International Conference on Language Re-sources and Evaluation (LREC), 2012.
4. A. Das, Björn Gambäck. Sentimantics: Conceptual Spaces for Lexical Sentiment Polarity Representation with Contextuality. In: Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis , 12 July 2012, pp. 38–46.
5. Namita Mittal, Basant Agarwal ,Garvit Chouhan , Nitin Bania ,Prateek Pareek. Sentiment Analysis of Hindi Review based on Negation and Discourse Relation. International Joint Conference on Natural Language Processing, October 2013,pp. 45–50.
6. Pooja Pandey, Sharvari Govilkar. A Framework for Sentiment Analysis in Hindi using HSWN. International Journal of Computer Applications(IJCA) (0975-8887). Vol.119-No.19, June 2015.
7. Shashank Sharma, PYKL Srinivas, Rakesh Chandra Balabantaray. Text Normalization of Code Mix and Sentiment Analysis. Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on. IEEE, 2015.
8. A. Joshi, A. R. Balamurali and P. Bhattacharyya. A Fallback Strategy for Sentiment Analysis in Hindi: a Case Study. In: Proceedings of the 8th ICON, 2010.
9. Balamurali A R, Aditya Joshi, Pushpak Bhattacharyya. Cross-Lingual Sentiment Analysis for Indian Languages using Linked WordNets. In: Proceedings of COLING 2012: Posters, pp 73–82.
10. D. Sai Krishna, G Akshay Kulkarni,A. Mohan. Sentiment Analysis-Time Variant Analytics. IJARCSSE, 2015, pp. 466-472.
11. Kamal Sarkar, Saikat Chakraborty. A sentiment analysis system for Indian language tweets. Springer international Publishing .2015 , pp. 694-702.
12. Aspect based Sentiment Analysis in Hindi: Resource Creation and Evaluation. Md Shad Akhtar, Asif Ekbal and Pushpak Bhattacharyya. International Conference on Language Re-sources and Evaluation (LRTC) .2015.
13. Mukesh Yadav, Varunakshi Bhojane. Sentiment Analysis on Hindi Content: A Survey. International Journal of Innovations & Advancement in Computer Science (IJIACS). December 2015 , ISSN 2347 – 8616 Volume 4, Issue 12.
14. A Das, S Bandyopadhyay. Phrase-level Polarity Identification for Bangla. International Journal of Computational Linguistics and Applications (IJCLA),2010, Vol. 1, No. 1-2, pp. 169-182.
15. Manju Venugopalan , Deepa Gupta. An Enhanced Polarity Lexicon by Learning-based Method Using Related Domain Knowledge. International Journal of Information Processing and Management (IJIPM), 2015, Vol. 6, No. 2, pp. 61 – 72.

.