www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 3 Issue 5 May, 2014 Page No. 5873-5878

ASPECT EXTRACTION & SEGMENTATION IN OPINION MINING

Mily Lal[1], *Kavita Asnani*[2]

[1]ISB&M School of Technology, Pune University
milylike@gmail.com

[2]Padre Conceicao College of Engineering, Goa University,
kavitapcce@gmail.com

Abstract: ***Opinion mining or sentiment analysis is the computational study of people's opinions, appraisals, attitudes, and emotions towards their aspects. Opinion mining is one of the significant areas of research in Web mining and Natural Language Processing (NLP) recently. With the growth of e-commerce, web documents are increasingly being used for decision making by individuals and organizations. This paper focuses on the identification of aspects related to customer opinions. Most of the recent works concentrate on explicit aspects only. Very few of them have dealt with implicit aspects. Here both explicit as well as implicit aspect is considered. A multi aspect review sentence will be segmented into multiple single aspects by segmentation because different opinions can be expressed on multiple aspects simultaneously in the same review. Opinions are polled to determine positive or negative comments for efficient decision making.***

Keywords: aspect, aspect extraction, opinion mining, opinion word, opinion summary, segmentation, Sentiment analysis

1. Introduction

Opinion mining and sentiment analysis has increased in recent years as a consequence of the growing availability and popularity of ecommerce. Opinion mining helps people seeking information to obtain and understand what others think. Traditional information extraction techniques cannot be applied effectively to opinion mining or sentiment analysis. It requires specialized techniques because of the unstructured nature of text in customer reviews and the huge amount of customer reviews.

Aspect based opinion mining deals with the extraction of opinion aspects by applying various techniques [3]. Aspect is defined as subject of the review such as the product and its specific attributes like "phone" ,"display", "screen" etc. Aspects can be divided in to explicit and implicit. By explicit it means that the aspect itself appears in the text, for example in the following sentences:

"The display is clear."

"The phone is small enough to put in my pocket."

The aspect "display" appears in the first sentence explicitly. However, in the second sentence, the aspect "size" is not directly mentioned but only implied by the word "small".

The main tasks of opinion mining are 1) identification of aspects 2) classification of aspect opinions into positive and negative and 3) Summarization of the results. Aspect identification is still a challenging task. Most recent works concentrated in the extraction of explicit aspects [2][3]. This paper focuses on implicit aspect extraction. It also focuses on segmentation of multi aspect review sentence into multiple single aspects because people often express different opinion on multiple aspects simultaneously on the same sentence. Finally poll the sentiments to find out customer satisfaction by determining positive and negative comments. Summarized results are presented to the users for their analysis.

This paper proposes many techniques based on data mining techniques and natural language processing for determining product aspects and its opinions. The techniques used here were evaluated to be quite effective in achieving the goals.

2. Related Works

This work is mainly related to four areas of research i) Aspect extraction ii) Aspect reduction iii) Association rule mining iv) Summarization.

2.1 Aspect Extraction

The majority of aspect extraction techniques fall into the following categories: language rule mining, sequence models and topic modeling.

Hu and Liu (2004) [3] first proposed a technique to extract product aspect based on association rule mining. Consumers tend to use the same words when they comment on the same product aspects, and then frequent itemsets of nouns in reviews are likely to be product aspects while the infrequent ones are less likely to be product aspects. Zhuang et al. (2006) [10] first identified reliable dependency relation templates from training data, and then used them to identify valid aspect-opinion pairs in test data. *Double propagation* (Qiu et al., 2011) [11] further developed the idea. The method needs only an initial set of opinion word seeds as the input and no seed aspects are required. It is based on the observation that opinions almost always have targets, and there are natural relations connecting opinion words and targets in a sentence due to the fact that opinion words are used to modify targets. Double propagation works well for medium-size corpora. But, it may result in low precision and low recall for large and small corpora respectively. Jin et al. (2009a and 2009b) [13][14] utilized lexicalized HMM to extract product aspects and opinion expressions from reviews. Different from traditional HMM, they integrate linguistic features such as part-of-speech and lexical patterns into HMM. One limitation for HMM is that its assumptions may not adequately represent problems and lead to reduced performance. To address the limitation, Conditional Random fields (CRF) (Lafferty et al., 2001) [12] is proposed. It is an undirected sequence model and can introduce more features than HMM at each time step. Titov and McDonald (2008) [16] pointed that global topic models such as PLSA and LDA might not be suitable for detecting ratable aspects. Both PLSA and LDA use the bagof-words representation of documents, therefore they can only explore co-occurrences at the document level. Titov and McDonald (2008) [16] proposed multigrain topic models to discover local ratable aspects, which models two distinct types of topics: global topics and local topics.

2.2 Aspect Reduction

Pang Lee *et al.* (2002, 2004) [2] [8] used supervised learning in sentiment analysis with the aim of determining whether it could be treated as a special case of topic-based categorization with positive and negative topics. Naıve Bayes (NB), Maximum Entropy (ME), and Support Vector Machine (SVM) classifiers were tested to achieve this, with all performing well in topic-based categorization. Document words and symbols were used for features as either a unigram or a bigram bag-of-features. Unigram features performed better than bigram features. Feature Frequency (FF) and Feature Presence (FP) when tested revealed that by using a SVM with unigram FP better accuracy could be achieved (82.9%) in a 3-fold cross validation.

2.3 Rule Generation

By using the Apriori algorithm, Hu and Liu [3][5][25] generated all strong association rules to extract both implicit and explicit opinion features expressed in reviews. They ran Apriori algorithm on the transaction set of noun/noun phrases to generate frequent itemsets. After producing candidate features they applied compactness pruning and redundancy pruning to remove those features that are not genuine. Their method may be effective in identifying explicit features. However, a straightforward extension of their approach cannot be applied to a sentence comprising just one opinion word, eg., “heavy” or “expensive” , as no corresponding patterns (or rules) are generated. Furthermore, the quantitative results for implicit feature identification was not provided in [5][25].

2.4 Summarization

A simple way to use the results is to produce a *feature-based summary* of opinions on an object or multiple competing objects [3, 25]. Feature buzz summary shows the relative frequency of feature mentioned. Organizations can know what their customers really care about [28]. Object buzz summary shows the frequency of different products in competition. The popularity of different products or brands in the market place can be found out using this method. Changes of every aspect using trend tracking can viewed as time opinions during a particular time is recorded[28]. Finally, researchers have also studied the summarization of opinions by producing a short textual summary based on multiple reviews or even a single review [28].

3. Proposed Techniques

The Figure 1 gives an architectural overview of the proposed aspect segmentation and summarization system.
The system is divided into four main modules which perform the main tasks such as aspect extraction, aspect reduction, rule generation and summarization.

3.1 Aspect Extraction

Aspects mostly tend to be nouns/noun phrases and opinion words are mostly adjective modifiers, adverb modifiers. Thus they are assumed to be the same here. They are extracted with the help of Stanford POS tagger .Therefore it is domain independent and unsupervised, avoiding tedious and time-consuming work of labeling data for supervised learning methods. It works very well in medium–size corpus. But for large corpora, this method may result in extracting many nouns/noun phrases which are not product aspects. The precision of the method plummets. The reason is that during extraction, adjectives which are not opinionated will be extracted as opinion words, e.g., “*whole*” and “*recent*”. They are not opinion words but it can modify many kinds of nouns/noun phrases, thus leading to extracting wrong aspects. Iteratively, more and more noises may be introduced during the process.

The other problem is that for certain domains, some important aspects do not have opinion words modifying them. For example, in reviews of cell phone domain, a reviewer may say “*There is Bluetooth on my phone*”. “*Bluetooth*” is an aspect, but the word “*Bluetooth*” may not be described by any opinion adjective, especially for a small corpus. Every adjective, adverb and verb have some implicit polarity (positive, negative or neutral), associated with them. With this polarity they modify the orientation of the objects. Once the aspects are retrieved and also the modifiers with respect to these aspects, the next important task is to assign subjectivity scores to these modifiers. SentiWordNet is used to determine the polarity of each modifier.

SentiWordNet (SWN) is a lexical resource of sentiment information for terms in the English language introduced in [15] designed to assist in opinion mining tasks. Each synonymous set in SWN has a positive sentiment score, a negative sentiment score and an objectivity score. When the sum of these scores equals one, it indicates the relative strength of the positively, negativity and objectivity of each synonymous set. The drawback in using SWN is that it requires word sense disambiguation to find the correct sense of a word and its associated scores.

For example, *“an unpredictable plot in the movie”* is a positive phrase, while *“an unpredictable steering wheel”* is a negative one. There has been significant research into this problem, as it was out of scope to use any sophisticated word sense disambiguation for this paper, so simply highest positive and negative values were taken.

Using negating terms such as “not” and “no” play a great part in determining the orientation of a term. For example:· “The film was good.”, “The film was not good.” The term “good”, contained in both sentences carries positive connotation and a positive SentiWordNet score. The second sentence however

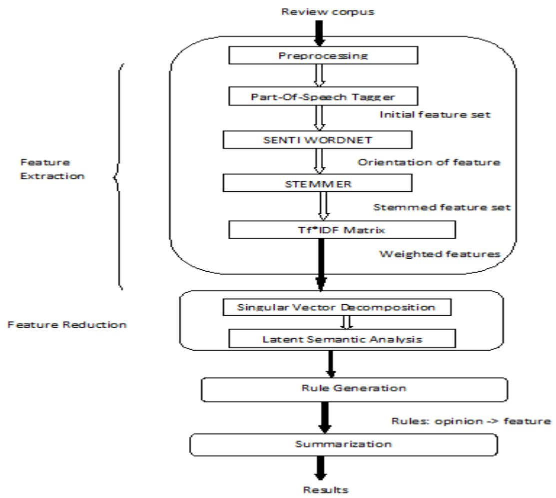

Fig. 1. Aspect Segmentation and Summarization system

has a negative meaning. In (Pang et al, 2002) [2] negation detection is modeled by adding a modifier prefix to negated terms, such as converting "great" into "not_great". The resulting modified text is then used as input for a word vector classifier.

To incorporate negation information, the negation word and negated word are joined with a hyphen. For example "*not good"* is replaced with *not-good*. The sentiment score of this new "word" is the negative of the sentiment score of the negated word [17]. In the English language, negation can occur in a variety of often subtle ways, therefore can lead to poor results.

It is worth highlighting here that, as noted in [18], lexical resources such as SentiWordNet are built independently of the data set being analyzed, and could be used in an unsupervised fashion, thus discarding the need for training data. The approach for a aspect set proposed in this section however starts from the principle that the aspects obtained through SentiWordNet capture diverse aspects of document sentiment, and are best suited for the creation of a data set that can be applied to train a classifier algorithm, like other machine n learning methods proposed in opinion mining.

3.2 Aspect Reduction

During sentiment analysis, a number of words which are used as aspects are considered, though only a few words in the corpus actually express sentiment. These extra aspects has to be eliminated as they slow down document classification as there are more words than really needed and secondly it reduces accuracy as the system should consider such words during processing. Using fewer aspects is advantageous and hence to remove those unnecessary aspects. As the name suggests, aspect reduction is the process wherein a corpus is run through to remove any unnecessary aspects. There is less information to consider leading to better accuracy.

Dimensionality Reduction is also beneficial since it tends to reduce *overfitting*, that is, the phenomenon by which a classifier is tuned also to the *contingent* characteristics of the training data rather than just the *constitutive* characteristics of the categories. Classifiers that over fit the training data is good at reclassifying the data they have been trained on, but much worse at classifying previously unseen data. Experiments have shown that, in order to avoid over fitting a number of training examples roughly proportional to the number of terms used is needed.

Latent Semantic Analysis is the proposed method used in this paper. It is one of the earliest approaches to reduce the dimension of vector representations of textual data (Landauer, McNamara, Dennis and Kintsch, 2007) [19].LSA was proposed as it is the only available method for text content based similarity deduction .Other methods do not rely on semantic representation but use semantic topic models based on generative models (e.g. probabilistic inference models like probabilistic latent semantic modelling and latent Dirichlet allocation [16].

LSA is a mathematical and statistical approach, claiming that semantic information can be derived from a word-document co-occurrence matrix and words and documents can be represented as points in a (high-dimensional) Euclidean space. Dimensionality reduction is an essential part of this derivation [20].

LSA requires relatively high computational performance and memory in comparison to other information retrieval techniques. Another challenge to LSA has been the alleged difficulty in determining the optimal number of dimensions to use for performing the SVD. Fewer dimensions allow for broader comparisons of the concepts contained in the document collection, while a higher number of dimensions enable more specific (or more relevant) concept comparisons. The actual number of dimensions that can be used is limited by the size and nature of the document collection.

3.3 Rule Generation

The proposed approach can be viewed as an elaborate extension of Hu and Liu's method [3][5][25]. Several differences are found. Firstly it is designed specifically to identify aspects that do not occur explicitly in review sentences. Secondly, the approach discriminates between opinion words and aspect words i.e opinion words can only occur in the rule antecedents, while rule consequents must be opinion aspects [26]. Thirdly association rules are generated directly from the LSA matrix of opinions and aspects.

Large number of incorrect rules may be generated which are caused by the incorrect identification of opinion words or explicit aspect words by the previous modules. However it helps in generating quite reasonable rules due to the LSA that helps to measure semantic associations between the objects [20].

3.4 Summarization

Different from traditional summarization, review summarization aims at producing a sentiment summary, which consists of sentences from a document that capture the author's opinion. The summary may be either a single paragraph as in [27] or a structured sentence list as in [3][10][25]. The former is produced by selecting some sentences or a whole paragraph which the author expresses his or her opinion(s). The latter is generated by the auto mined aspects that the author comments on. The proposed method used is more relevant to the method used in [3][25][10] i.e. *aspect based summary* of opinions on an object or multiple competing objects.

4. Implementation

Statistical Opinion Mining is used which tackles sentiment analysis in terms of data mining and is based on statistical methods. The product review corpus was collected from www.amazon.com. Amazon as the source of reviews, which includes user reviews for cell phones. The corpus contains 300 reviews. Products in this site have a large number of reviews. Each of the review includes a text review. Additional information available but not used in this project includes date, time, author name, location and ratings. Reviews based on cell phones were manually collected. A typical review contains free text summary about a product. All reviews are plain text.

First a set of standard preprocessing steps are carried out, viz., tokenizing and stemming. Stanford POS tagger was used for part of speech tagging which automatically classify words into categories based on parts of speech from the source documents. The following pattern was used to extract nouns and noun phrases from the Part Of Speech (POS) Tagger output:-i) NN or NNS ii) NN NN or NNS

Example:- In the sentence, "*This camera produces beautiful pictures*", "pictures" and "camera" will be extracted as it satisfies the first pattern. In the sentence, "*This is a simple cell phone"*, "cell phone" will be extracted as it satisfies the second pattern. be used with the SentiWordNet database. Positive and negative words are identified using SentiWordNet_3.0.0. Words with score less than 0 are taken as negative words and those with scores greater than 0 is taken as positive words. The remaining is considered neutral. Following Latent Semantic Analysis was performed. Values close to 1 represent very similar words while values close to 0 represent very dissimilar words. The terms with similarity score above a particular threshold is returned.

Association rule mining helped to mine the aspect->opinion rules from the LSA matrix. Support score threshold was taken to be 1% [3][5] so that weak rules are pruned appropriately. As the output of this stage a refined set of rules are returned.

After identifying all valid aspect-opinion pairs, the final summary is generated according to the following steps. First, all the sentences that express opinions on the queried aspect, are collected. Then, the semantic orientation of the relevant opinion in each sentence is identified from the sentiment score previously collected. Finally, the organized sentence list is shown as the summary [3][25][10].

In the case of implicit sentence, first a matched list of rules is collected by searching the above robust rules antecedents that are identical to the opinion word extracted. Due to the existence of synonyms and semantically related terms, it is better to associate an opinion word to most likely aspect cluster instead of a single aspect.

Example:-

"it's also a good tool for entertainment"

Opinions -> good

Implicit aspect = nokia-n95 -> [nokia-n95, memory, sound, memory-card]

Therefore among the matched robust rules, the one corresponding to the majority aspect cluster (containing the largest number of aspects) is then fired, and accordingly the representative word of the cluster is chosen as the identified implicit aspect [26]. If the implicit aspect is identical to the queried aspect then the sentence is organized in the summary else ignored.

5. Observations and Results

5.1 Aspect Extraction

Total # of words/ tokens generated from the reviews = 29844 words. The percentage reduction after stemming =25.6%

In Table1 the precision plummets as a result of extracting many nouns/noun phrases which are not product aspects.

In Table2 precision plummets as a result of extracting many nouns/noun phrases which are not product aspects. The reason is that during extraction, adjectives which are not opinionated will be extracted as opinion words, thus leading to extracting wrong aspects. Iteratively, more and more noises may be introduced during the process.

Table 1: Precision and Recall of aspects extracted by POS

No of Aspects Extracted by POS Tagger	**No of Aspects Correctly Extracted by POS Tagger**	**No of Aspects Not Extracted by POS Tagger**	**Precision**	**Recall**
313	210	41	67.09%	83.66%

Table 2: Precision and Recall of opinions extracted by POS

No of Opinions Extracted	**No of Opinions Correctly Extracted**	**No of Opinions Not Extracted**	**Precision**	**Recall**
188	136	68	72.34%	66.67 %

Table 3: Precision of Sentiment Polarity by SentiWordNet

Sentiment Polarity by SentiWordNet	Positive Opinions	107
	Negative Opinions	64
Sentiment Polarity Correctly identified by SentiWordNet	Positive Opinions	79
	Negative Opinions	37
Precision	Positive Opinions	73.83%
	Negative Opinions	57.81%

The cell phone data set used by Hu et al. (2004)[3] was used for testing. They manually read all the reviews. They produced a manual list of features.

Table 4: Precision aspects generated

No. of Manual aspects by Hu et al. (2004). [3]	**No of aspects correctly generated by POS**	**Precision**
111	92	82.88 %

Hu et al. (2004) [3] tagged all the aspects on which the reviewer has expressed his/her opinions. If the user gave no opinion in a sentence it was not tagged. Many aspects like "dependable ","compactable" were also found to be aspects. These words are extracted as sentiment words by the proposed model of the project.

Hu et al. (2004)[3] also identified negativity and positivity of the opinions extracted.

Table 5: Accuracy of Sentiment Orientation

Sentiment Orientation accuracy obtained by Hu, M. & Liu, B [3]	**Sentiment Orientation Accuracy by SentiWordNet**
76.4%	65%

Table 5 result shows the shortcoming of the method used for assigning sentiment polarity. Assigning the same polarities of the opinion words to the product aspects does not work well in most situations. Hu et al. (2004). [3] After extracting the potential opinion words, they identified the polarities of the opinion words by utilizing synonymous set and antonymous set in the WordNet, and a small list of opinion words with opinion polarities

5.2 Aspect Reduction

LSA has been the alleged difficulty in determining the optimal number of dimensions to use for performing the SVD k was determined to be "6" as the optimal number of dimensions to use for performing the SVD as recall as 66.67% and precision as 71.42%.

Table 6: Precision and Recall of LSA

K value	precision	recall
5	75.62	68.5
6	71.42	66.67
7	69.32	64.13

Too few dimensions and important patterns are left out, too many and noise caused by random word choices will creep back in. The table below shows the advantage of conceptual mapping for retrieving implicit aspects**.**

Table 7: Precision and Recall of LSA for retrieving implicit aspects

K value	Mapping Analytically		Mapping conceptually	
	precision	recall	precision	recall
5	15.78	17.79	54.88	61.86
6	16.56	17.94	62.72	67.94
7	17.48	19.30	63.22	69.80

Due to the existence of synonyms and semantically related terms, associating a opinion word to most likely feature cluster instead of a single aspect was found better. An aspect cluster contains features that are conceptually and semantically related. The representative word of the cluster is chosen as the identified implicit aspect. Mapping of a single aspect failed in many situations.

Example:-For every occurrence of opinion word "horrible" in an implicit sentence , the aspect "speaker " will be returned as it is the representative aspect of the cluster which contains all the features associated with the opinion word "horrible". In the sentence "It is the most horrible thing I have ever owned". Here "horrible" might be referring to aspects like "phone", "camera" etc. In such situation the aspect cluster might contain the required aspect as given.

Cluster of "speaker" contains "music-system, music, phone, service, music-player". Therefore conceptual mapping of aspects with opinions helps to improve the performance. This shows strength of LSA usability being much more than the dimensionality reduction.

5.3 Rule Generation

Table 6: Precision and Recall of Rules Mined

Total No Rules Mined	Correctly Mined Rules	No of all Correct rules	Precision	Recall
943	494	688	53.57	71.8

1% was taken as minimum support for generating association rules. Incorrect rules can be generated due to the incorrect identification of opinion words or explicit aspect words by the previous modules. Quite number of reasonable rules was generated due to the LSA that helps to measure semantic associations between the objects. Higher minimum support will weed out many lower frequency associations but will kill the legitimate associations as well. Therefore Precision values were observed to be lower than Hu.(2004)[3] which obtained a precision of 68.65 and Popescu(2005) [4] got 59.65 % .The recall values were observed to be quite satisfying as it is better than Hu.(2004)[3] with 57.93% and Popescu(2005) [4] with 59.95%.

6. Conclusion

This paper proposes a set of techniques for extracting and summarizing product reviews based on data mining and natural language processing methods. The objective is to perform aspect extraction and segmentation which is an essential component of an aspect-based opinion mining system. Without knowing the aspects, the mined opinions are of little use.

For aspect and opinion extraction, POS tagger is being used. Sentiment orientation is derived with the help of SentiWordNet. Latent semantic Analysis is used for aspect reduction. It performs similarity deduction which is useful in deriving implicit aspects by conceptual mapping of opinions with aspects. Opinion-aspect mapping is done using Apriori Algorithm. It is used for finding opinions and aspects pairs which frequently occur together. The opinion-aspect pairs thus generated are used for identifying implicit aspects. Finally aspect based summary of the customer reviews a product sold online is provided.

Although the methodology used here is quite reasonable in identifying implicit and explicit aspects some undesirable errors still exist in the final identification results. Some of the errors are caused by incorrect identification of explicit aspects and opinion words, while other errors caused during segmentation or parsing.

7. References

[1] Turney, p. (2002). Thumbs up or thumbs down? Semantic orientation applied tounsupervised classification of reviews, In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Pennsylvania.

[2] Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification Using Machine Learning Techniques, In Proc. of EMNLP.

[3] Hu, M. and Liu, B. 2004. Mining and summarizing customer reviews. International Conference on Knowledge Discovery and Data Mining (ICDM).

[4] Popescu, A. and Etzioni, O. (2005). Extracting product features and opinions from reviews, In Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP). Potts, C. (2007) The Expressive Dimension. Theoretical Linguistics 33:165-198.

[5] Ding, X., Liu, B. and Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining, In Proceedings of the Conference on Web Search and Web Data Mining (WSDM).

[6]Qi Su,Kun Xiang,Houfeng Wang,Bin Sun and Shiwen Yu(2006).Using Pointwise Mutual Information to Identify Implicit Features in Customer Reviews.ICCPOL ,LNAI 4285 ,pp.22-30,Springer(2006).

[7]Zhu J.,H.Wang,M,Zhu and B.K.Tsou.2011.Aspect based opinion polling from customer reviews. IEEE Transactions on Affective Computing,2(1):37-49.

[8] Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of ACL-04, 42nd Meeting of the Association for Computational Linguistics, pages 271–278, Barcelona, ES.

[9] Isabella, J **"**Analysis and evaluation of Feature selectors in opinion mining", Indian Journal of Computer Science and Engineering (IJCSE), Vol. 3 No.6 Dec 2012-Jan 2013

[10] Zhuang, L., F. Jing, X.-Yan Zhu, and L. Zhang. Movie review mining and summarization. In Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2006), 2006.

[11] Qiu, G., B. Liu, J. Bu, and C. Chen. Opinion word expansion and target extraction through double propagation. Computational Linguistics, 2011.

[12] Lafferty, J., A. McCallum, and F. Pereira. Conditional random fields: probabilisticmodels for segmenting and labeling sequence data. In Proceedings of International Conferenceon Machine Learning (ICML-2001), 2001.

[13] Jin, W. and H. Ho. A novel lexicalized HMM-based learning framework for web opinion mining. In Proceedings of International Conference on Machine Learning (ICML-2009), 2009a.

[14] Jin, W. and H. Ho. OpinionMiner: a novel machine learning system for web opinion mining and extraction. In Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2009), 2009b.

[15] Esuli, A., and F. Sebastiani. "Determining term subjectivity and term orientation for opinion mining." In Proceedings of Annual Conference of the European Chapter of the Association of Computational Linguistics (EACL-2006), 2006.

[16] Titov, I. and R. McDonald. "Modeling online reviews with multi-grain topic models." In Proceedings of International Conference on World Wide Web (WWW-2008), 2008.

[17] Shitanshu Verma, Pushpak Bhattacharyya."Incorporating Semantic Knowledge for Sentiment Analysis". Proceedings of ICON-2008: 6th International Conference on Natural Language Processing Macmillan Publishers, India.

[18] Pang B, Lee L. (2008) "Opinion Mining and Sentiment Analysis", Foundations and

Trends in Information Retrieval, Vol. 2, Nos. 1-2, pp. 1-135, 2008.

[19] Ladauer, T.K., McNamara, D.S., Dennis, S., & Kintsch,W.(Eds). (2007) Handbook of latent semantic analysis. Mahwah, NJ:Lawrence Erlbaun.

[20] Deerwester, Dumais, Furnas, Lanouauer, and Harshman,Indexing by latent semantic analysis, Journal of the American Society for Information Science, 41 (1990), pp. 391-407.

[21] J. Wiebe and R. Mihalcea. "Word sense and subjectivity." Proceedings of the Conference on Computational Linguistics / Association for Computational Linguistics (COLING/ACL), 2006

[22] J. Wiebe, "Learning subjective adjectives from corpora,". Proceedings of AAAI, 2000.

[23] Bing Liu. Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012.

[24] Kristina Toutanova, Dan Klein, Christopher Manning, and Yoram Singer. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In Proceedings of HLT-NAACL 2003, pp. 252-259.

[25] Liu, B., Hu, M., And Cheng, J. 2005. Opinion observer: analyzing and comparing opinions on the web. In WWW '05: Proceedings of the 14th international conference on World Wide Web. ACM, New York, NY, USA, 342–351.

[26] Zhen Hai, Kuiyu Chang and Jung-jae Kim. Implicit Feature Identification via Co-occurrence Association rule Mining.Springer(2011)

[27] Akshat Bakliwal,Piyush Arora , Vasudeva Varma. Entity Centric Opinion Mining from Blogs. Preceedings of the 2nd Workshop on Sentiment Analysis where AI meets Psychology(SAAIP 2012),pages 53-64, COLING 2012,December 2012

[28] Liu, B., "Sentiment Analysis and Subjectivity" Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau), 2010

Author Profile

Mily Lal received B.Tech-Computers degree from Calicut University and M.E-I.T degree from Goa University. She is Assistant Professor of Computer department in ISB&M School of Technology,Pune

Kavita Asnani received B.E-Computers degree from Shivaji University and M.E-I.T degree from Goa University. She is now HOD incharge of I.T department in Padre Conciecao college of Engineering, Verna,Goa