computing@computingonline.net
www.computingonline.net

Print ISSN 1727-6209
On-line ISSN 2312-5381
International Journal of Computing

OPINION TARGET EXTRACTION WITH SENTIMENT ANALYSIS

Surbhi Bhatia [1], Manisha Sharma [2], Komal Kumar Bhatia [3], Pragyaditya Das [4]

[1] CS Department, Banasthali University, India, surbhibhatia1988@yahoo.com,
[2] CS Department, Banasthali Vidyapith, India, manishasharma8@gmail.com
[3] CSE Department, YMCA University of Science & Technology, India, komal_bhatia1@rediffmail.com
[4] ICE Department, National Institute of Technology Tiruchirappalli, Tamil Nadu, India

Paper history:
Received 23 March 2018
Received in revised form 31 July 2018
Accepted 28 August 2018
Available online 30 September 2018

Keywords:
Opinion mining;
Ontology;
Information extraction;
Sentiment Analysis;
Supervised Learning.

Abstract: Social networks have increased their demand extensively for mining texts. Opinions are used to express views and reviews are used to provide information about how a product is perceived. The reviews available online can be available in thousands, so making the right decision to select a product becomes a very tedious task. Several research works has been proposed in the past but they were limited to certain issues discussed in this paper. A dynamic system is proposed based on the features using ontology followed with classification. Classifying information from such text is highly challenging. We propose a novel method of extracting aspects using ontology and further categorizing these sentiments into positive, negative and neutral category using supervised leaning technique. Opinion Mining is a natural language processing task that mine information from various text forums and classify them on the basis of their polarity as positive, negative or neutral. In this paper, we demonstrate machine learning algorithms using WEKA tool and efficiency is evaluated using information retrieval search strategies.

1. INTRODUCTION

Internet users post their comments and views on the various social net-working sites such as Twitter, Amazon, Facebook, Flip kart etc. The text present in these sites are of unstructured nature and has been immensely increasing as people engage themselves more into Social media. Instead of surveying about the market explicitly, the business intelligence and the demand in market related to a particular product can be achieved by the reviews posted online. The text information present needs to be converted into structured data for sentiment analysis. Sentiment analysis is a synonym used for opinion mining as it determines the contextual polarity of a document [1]. Sentiment analysis has been divided into three classification levels: document, sentence, and aspect [2]. Document level sentiment classification deals with extracting words containing opinions and then determining polarities of these words from the reviews posted. [3,4]. Sentence-level sentiment classification deals with first identifying whether a sentence is objective or subjective and then classifying these small sentences into positive or negative category [5]. The finest method is Aspect/feature-level Sentiment analysis which extracts features from opinions and then on the basis of features, grouping them into classes. Thereafter, the polarity is determined and the summarized results are shown as final output [6].

Opinions can be represented as: (target, sentimentvalue, holder, time). The target can be a single entity or a combination of many entities. According to Liu [7, 8], description of an entity can vary from an item, individual, society, or subject. It can be represented as a chain of components and its sub parts.

The target is broken down into aspects and entity. An opinion is a quintuple (targetentity, aspect/feature, sentimentvalue, holder, time) as suggested by [9]. In the above example, Samsung Tablet is the target entity, picture quality and battery are the aspects, positive is the sentiment value, 'I' is the holder; 'dd-mm-yy' is the time when the review was posted. Objective sentences are facts whereas subjective information is views or opinions expressed as unstructured text [10]. We work on finding the explicit aspects using ontology. We can define Ontology as official and obvious requirement of a shared conceptualization [11]. We work on identifying aspects both explicit and implicit by using dependency relations and concepts, i.e. using ontologies and implicit aspects by manually developing an implicit repository for respective domain. The sentiments are searched within sentences containing re-views by using a wordlist and within those searched words will serve our purpose of getting the implicit features.

Thereafter, we work on categorizing the sentences into positive or negative sentiment in a novel way. The algorithm proposed in totally supervised and is based on the training the data. An in depth comparison of the classifiers has been carried out using Weka tool [12]. The classifiers are examined in terms of the error rate and accuracy. Various information retrieval search strategies are evaluated and illustrations are shown in the form of graphs and charts. The classifiers used are Naive Bayes, Logistic Regression and Decision trees.

The structure of the paper is described as follows. Section 2 discusses the existing work done in the area, section 3 presents the concept of Aspect extraction, proposes the novel method of extracting both explicit and implicit features. Section 4 explains all the modules sequentially. Section 5 and 6 discusses the implementation of our proposed work and presents the analysis by evaluating the work. Section 7 concludes.

2. LITERATURE REVIEW

The Opinion Mining is the field of Data mining which helps in determining semantic orientation of the overall text by making use of expressions as a “bag of sentiment words” and assigning values to those expressions as positive, negative or neutral towards a given subject [13]. The paper focuses on implicit feature detection by considering a text classification problem by using centroid based approaches. Although it performs better than rule based approaches, but still incorrect classification can make wrong identification of implicit feature words [14]. The paper used ontologies where both implicit and explicit features are identified by finding relationships between concepts and lexical information. Lot of semantic information is required in building ontologies [15]. The paper identified only explicit aspects for accommodation domain based on frequency and position in texts by dividing reviews in three equal parts. No work on preprocessing tasks like lemmatization is done; also the work is limited to a single domain [16]. The paper has determined explicit domain specific generic features by extracting nouns representing features and eliminating nouns that do not represent features by using association mining and probabilistic methods. Summarization is also achieved using clustering and evaluation is illustrated on the basis of information retrieval search strategies [17]. This paper used unsupervised approach by proposing a novel graph based algorithm for domain specific ontology extraction [18]. The paper discusses the classification of sentiments using Naive Bayes method. The classifier is implemented in Python using hash tables and results have been evaluated [19].

Orimaye et al. detailed out the tasks of sentiment analysis and showed comparison of various methods [20]. Samha et al. focused on detecting aspects by creating it manually. Related aspects from reviews were determined from the online lexicon; with the help of WordNet [21]. The paper has explored all the Potential implicit features of restaurant reviews and product reviews by assigning score and calculating frequency of the sentence considering similarity measure between words of that sentence. After passing from the threshold, the one with the highest-scoring will be assigned to that sentence [22].

3. ASPECT EXTRACTION

The Aspect Extraction involves extracting the features from the opinion. We will concentrate on two modules in aspect extraction. Target entity extraction module and Opinion word lexicon module.

The former deals with extracting the entity, aspect/features from the structured review and the latter deals with classifying it into positive, negative and neutral category.

3.1 TARGET ENTITY EXTRACTION:

Our technique is a novel based which works on extraction of explicit aspects. Since we are working on Aspect based extraction, so extracting features is our prime objective. Features are of two types, Implicit and explicit. Researchers have studied widely in the field of explicit aspects and many methods have been estimated for the same. However, very little work has focused on the identification of implicit aspects due the complexity of tracing them from reviews [23]. The proposed technique works on extracting explicit and implicit aspects by developing rules.

For the sentences that contain nouns:

a) For the sentences containing nouns with adjectives

For explicit aspect extraction, we have extracted the target entities using manually built ontology following the transitive rule If A->B and B-> C, the A->C; where A is related to the domain we are developing the ontology, B are the relations and C are the methods, attributes, synonyms etc. Earlier research on opinion indicates that the features of a product are usually represented using nouns [24]. We will extract the explicit features based on the earlier research with some modification. The nouns are extracted and stored in the repository. The most prominent words representing features are also extracted using ontology matrix and stored in a separate file. Now the rules are formed

The words which are there in both the lists are taken directly.

The words which are in of the lists are taken and count is applied. If the count falls below the threshold set (by conducting various experiments on different datasets), then the word is ignored, else taken as representing feature. After extracting the opinion words using nouns or using ontology, the features are extracted by computing the distance of each opinion word using K means clustering to detected opinion words in a sentence and then assigning an opinion to the feature. Opinion word for the feature will always be the closest one around it. Most researchers have proved the fact that if frequent aspect is missing in a sentence containing opinion word, then the closest noun is considered and paired with the opinion word. On looking into the structure of the ontology, we can remove the limitation as there may be nouns which may not represent features. In the research done by Bafna, for the words that do not represent features but they are nouns, will be ignored in our work because the threshold will fall below the finalised limit, as we are commenting on the opinion in the example: my comments are always accurate with respect to iPhone reviews" is not a review. In those cases, ontology matrix will serve our purpose. Also a threshold is set by conducting various experiments on different datasets to filter out the opinion words using ontology. The paper overcomes the problem of frequent aspect extraction as not all the aspects that are frequent as infrequent aspect as reported by Hu and Liu [25].

But for other sentences, implicit features come into play. We have created our own implicit dictionary. It will work with supervised learning approach. We cluster it with the related category as proposed by [26]. The list keeps on updated by checking if the words can be put into the related category by checking for their synonyms and their antonyms.

For example: Taking the example of laptop, we can have implicit repository with the following attributes of phone.

long->length->life->battery->battery life
price->cost->expensive
weight->size->heavy>carry
wide->large->big>hold>screen
sound->noise->interacts->talk->voice

b) For the sentences that contain nouns with verbs:

The laptop lasts long (refers to battery)
Big to hold (refers to screen)

For the sentences that do not contain nouns:

a) Verbs with adjectives:

Interacts in a beautiful way (refers to sound)

b) Adverb with adjectives

To carry longer in hand is tedious (refers to weight)

The opinion feature pairs are identified and stored in the Opinion feature pair repository for further processing. The detailed process is shown in Fig 1.

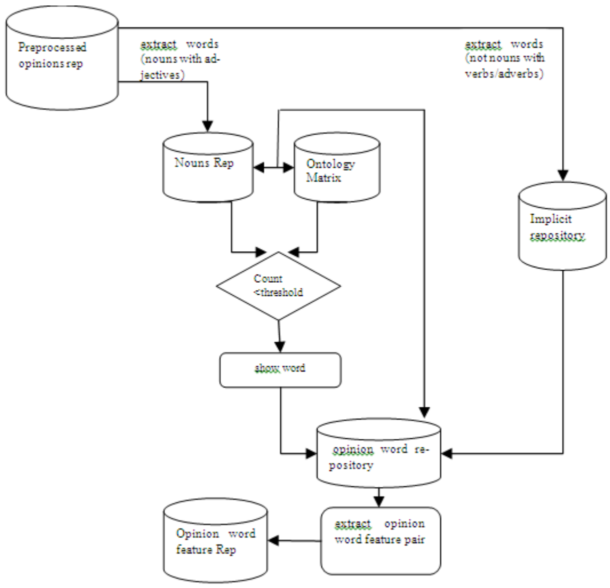

Figure 1 –Detailed Architecture

Generally, the extracted features correspond exclusively to terms contained in the ontology. The ontology matrix for laptop has been constructed manually [27]. The ontology matrix for Sony Laptop has been constructed manually as shown in Table 1.

Table 1. Ontology Matrix

Laptop audio	Has a	Microphone
Laptop audio	Has a	Stereo Speakers
Dimensions	Has a	Laptop height
Dimensions	Has a	Laptop width
Display Size	Is a	13 inch
Display Size	Is a	15 inch
Display Size	Is a	17 inch
RAM Memory	Is a	4GB
RAM Memory	Is a	8GB
OS	Type of	Windows
OS	Type of	Linux
Processor	Has a	IntelCore I3
Processor	Has a	IntelCore I5
Processor	Has a	IntelCore I7
Storage	Type of	HDD
Wireless	Type of	Bluetooth
Wireless	Type of	WiFi
Laptop	Feature of	Battery
Laptop	Feature of	Laptop Audio
Laptop	Feature of	Laptop Brand
Laptop	Feature of	Laptop Camera
Laptop	Feature of	Dimensions
Laptop	Feature of	Memory
Laptop	Feature of	Display size
Laptop	Feature of	OS
Laptop	Feature of	Processor
Memory	Is a	RAM
Memory	Is a	ROM
Laptop	Feature of	Storage
Laptop	Feature of	Wireless
Laptop	Action on	Turn
Laptop	Action on	Connect
Laptop	Action on	Restart
Laptop	Action on	Identify
Laptop	Action on	Remove
Laptop	Action on	Lock
Laptop	Action on	Use

3.2 OPINION WORD LEXICON:

We classify the opinions into positive, negative or neutral category by assigning the polarity scores to each token. Each token is checked with its dictionary meaning and thereafter, SentiWordNet make our method of assigning scores unique and more efficient. Technically, the resource contains Princeton WordNet data marked with polarity scores [29]. They particularly assign polarity scores to each Sysnet in the WordNet [28]. Use of dictionary will help in finding synonyms and antonyms relations of the words which are not present in the Opinion lexicon for determining the polarity of new opinion words [30]. These words will be added to our database.

After checking the token with the dictionary meaning and SentiWordNet, the final scores are given to each token for summarization.

The summarization is done by producing a cumulative one liner summary for the product. This is done by adding all the positives and negatives and finding whether positivity count is greater or negativity count is more. If both counts appear to be equal, we consider it as neutral score.

4. PROPOSED WORK

Opinions Extractor:

Reviews are extracted either using the crawler developed [31] or parsing the HTML code. The tweets are taken from www. Twitter.com Using the HTML part of tweets and stored in the tweets repository.

Preprocessing:

After extracting these tweets, the pre-processing tasks like lemmatization, stop word removal, removing punctuations and special characters have been performed. The preprocessed repository contains the filtered tweets

POS Tagger:

Stanford POS tagger has been used to split the tweets into tokens. All the tokens are stored with its appropriate part of speech in the Tweets Tokens Repository. The part-of-speech tagging will categorize the English grammar in nouns, verbs, adjectives, pronouns, prepositions, conjunctions and interjections. For POS tagging the documents, we used Stanford NLP Parser [32].

After text processing, the aspects are extracted and tweets are classified into positive, negative category and finally a summarized view is given as recommendation or rejection of the product.

5. IMPLEMENTATION

The algorithm has been implemented in Python using R language.

The extracted tweets from Twitter.com are preprocessed removing all the special characters, alphanumeric characters. The tweets contain words such as true, false. These words are replaced with neutral words such as 'blablaaa'. Then the features are extracted using the above approach and tweets are classified into positive, negative and neutral category.

5.1 SNAPSHOTS:

We conducted some initial tests of proposed approach. We took in total of 450 Sony laptop reviews and 360 lens reviews. The polarity was predicted into positive score and negative score. We evaluated the algorithm against the human evaluation. We also found that battery, applications were most prominent in the laptop reviews. And the concepts lens, glass and cases were the most mentioned in the lens re-views.

Other concepts were not explicitly mentioned in the reviews therefore for implicit reviews, we worked on their synonyms. Implementation details include the use of Weka tool, version 3.8 on the Ubuntu 16.04.2 platform that was installed using Oracle VM virtual Box 5.1.18. The results are shown in Fig 2.

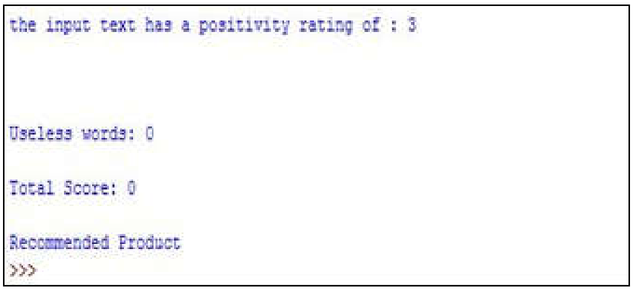

Figure 2 –Final Scoring

Step 1: Fetching tweets with #Sony Laptop tags. (R)

Storing the fetched data in a text file named Tweets.txt

Next, we do the processing:

Step2 : Preprocessing

Removing special characters.

Removing all FALSE-NEGATIVE and FALSE POSITIVES. (Replacing words like TRUE, FALSE etc.(which are the part of the HTML attribute of the Tweet) with neutral words).

Step 3: Aspect Extraction and Scoring the tweets: (PYTHON)

Text mining and scoring:

6. EVALUATION

The three basic performance metrics taken in consideration for evaluating the proposed work are as follows:

Precision: The ratio of the appropriately categorized tweets over all the tweets by the proposed algorithm. (Correctly crawled opinions and incorrectly crawled opinions). Mathematically, the Precision is given by: $P = OT / (OT+WTP)$, where OT is Relevant tweets retrieved and WTP is the number of Irrelevant tweets retrieved. It is usually expressed as a percentage.

Recall: The ratio of the appropriately categorized tweets by the proposed technique over all the tweets as given by the experts. Mathematically, the Recall is given by: R = OT/ (OT + NTP), where OT Relevant tweets retrieved and NTP Relevant tweets not retrieved. It is usually expressed as a percentage.

F-Measure: The combination of the above explained two values is F-measure. F = 2PR/ (P + R), where Precision P and Recall R are equally weighted.

6.1 ANALYSIS:

The complete dataset prepared in text file is exported to csw file after classifying the sentiments and conversion is done in the arff format (attribute relation file format), since analysis of our work is demonstrated using WEKA tool. We conducted analysis for the healthcare and electronics domain. The analysis for the work is carried forward and the accuracy is calculated using commonly used supervised learning algorithms. These algorithms works well even with less training data and its easy to understand the results. The effectiveness of the algorithms is calculated using standard information retrieval parameters discussed in the above section [33]. The error rate is depicted using various graphs and charts.

The analysis of our work has been shown using different algorithms using the data set "laptop.arff" in Table 2.

Combining training and test data, we have accumulated 435 instances for analysis. The data consists of 17 attributes. The performance of the classifiers is shown below. We have used various algorithms like Naive Bayes, Decision trees, Logistic Regression for analysis and compared their efficiency in terms of error rate, accuracy and time to build the model.

Table 2. Laptop.arff

```
@relation laptop

@attribute processor_i {1, 0}
@attribute display-greater-than-14 {1, 0}
@attribute memory-greater-than-2gb {1, 0}
@attribute hdd-greater-than-1tb {1, 0}
@attribute os-windows {1, 0}
@attribute high_audio {1, 0}
@attribute graphics_card {1, 0}
@attribute camera_hd {1, 0}
@attribute wireless {1, 0}
@attribute bluetooth {1, 0}
@attribute weight-less-than-4lb {1, 0}
@attribute color-black {1, 0}
@attribute brand-dell {1, 0}
@attribute memory-type-ddr3 {1, 0}
@attribute processor-brand-intel {1, 0}
@attribute series-inspiron {1, 0}
@attribute Class {1, 0}

@data
'0','1','0','1','1','1','0','0','0','1','1','1','1','1','0','1','1'
'0','1','0','1','1','1','0','0','0','0','0','1','1','1','0','1','1'
'1','1','1','1','1','1','0','0','0','0','1','0','1','1','0','0','0'
'0','1','1','0','1','1','0','0','0','0','1','0','1','0','0','1','0'
'1','1','1','0','1','1','0','0','0','0','1','1','1','1','1','1','0'
'0','1','1','0','1','1','0','0','0','0','0','0','1','1','1','1','0'
'0','1','0','1','1','1','0','0','0','0','0','0','1','1','1','1','0'
'0','1','0','1','1','1','0','0','0','0','0','0','1','1','1','1','1'
'0','1','0','1','1','1','0','0','0','0','0','1','1','1','0','1','1'
'1','1','1','0','0','0','1','1','1','0','0','0','0','0','1','1','0'
'0','1','0','1','1','0','0','0','0','0','1','1','1','1','0','0','1'
'0','1','0','1','1','1','0','0','0','0','1','1','1','1','1','1','1'
'0','1','1','0','0','0','1','1','1','0','0','0','1','0','1','1','0'
'1','1','1','0','0','1','1','1','1','1','1','1','0','0','1','1','0'
'0','1','0','1','1','1','0','0','0','0','0','1','1','1','0','1','1'
```

Cross-validation is defined as a way of making inference and assessing how a dataset will perform on an indefinite dataset [34]. It is a technique of generalizing the model. We evaluated three different supervised learning algorithms and validation was achieved by using three fold cross validation. We divided the dataset into three sections in which two sections were used as a training set and one as a test set and experiments were conducted and results of cross validation are shown by applying Word Tokenizer.

Cross validation results for all the three algorithms are shown in Table 3, 4 and 5.

Table 3. Naive Bayes Results

Naive Bayes		Predicted Class		
Actual Class		Negative	Positive	Total
	Negative	155	13	168
	Positive	29	238	267
	Total	184	251	435

Table 4. J-48 Results

J-48		Predicted Class		
Actual Class		Negative	Positive	Total
	Negative	161	7	168
	Positive	9	258	267
	Total	170	265	435

Table 5. Logistic Regression Results

Logistic Regression		Predicted Class		
Actual Class		Negative	Positive	Total
	Negative	156	12	168
	Positive	15	252	267
	Total	171	264	435

6.2 RESULTS:

The performance of the techniques is evaluated by calculating the accuracy and error rate by using the given formula below:

Accuracy = Number of True Outcomes/Total Number of Predictions

Error Rate = Number of False Outcomes/Total Number of Predictions

On calculating the accuracy, we received the following results shown in Table 6 for laptop and lens respectively.

Table 6. Accuracy results

Supervised Learning Algorithms	Accuracy(%)		Time to build the model (sec)		Features extracted (tokens [per word])		Error Rate (units)	
	laptop	Lens	Laptop	Lens	Laptop	Lens	Laptop	Lens
Logistic Regression	93.7	83.3	0.26	0.07	740	358	0.06	0.11
Naive Bayes	90.3	83.3	0.03	0.01	740	358	0.09	0.24
J-48	96.3	87.5	0.18	0.02	740	358	0.06	0.15

Finally, it is worth noting that precision and recall scores of the proposed work outperforms results of other groups as mentioned in section 2 by a good margin. The error rate has been improvised to 12% with the proposed technique. The analysis is shown in chart form for laptop in Fig 3.

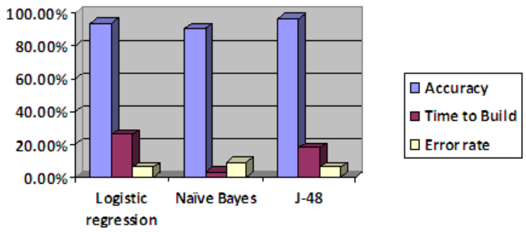

Figure 3 –Analysis

7. CONCLUSIONS AND FUTURE TRENDS

In this paper, identification of opinions is carried out on the basis of ontologies. Our work has focused on extracting explicit and implicit aspects using ontologies and rules. The work has been proposed using a novel technique and has been implemented taking the real set data in the form of tweets. The analysis has been performed and the three-fold cross validation results were used to evaluate the algorithms like Naive Bayes, Logistic regression and J-48. Our future work will focus on extracting implicit aspects incorporating some rules of Natural Language Processing and by improvising hybrid feature selection methods by properly inserting formulas as equation in the text. Domain ontology and sentiment lexicon were needed as pre requirements and final polarity orientation task is achieved showing the analysis in Weka tool. According to these results, applying the algorithm we find that for J-48 algorithm, precision for laptop concepts with positive polarity of 0.96 for lens comments we found f-measure positive polarity of 0.87.

8. REFERENCES

[1] S. Bhatia, M. Sharma, K. K. Bhatia, "Strategies for mining opinions: A survey," Proceedings of the IEEE 2nd International Conference on Computing for Sustainable Global Development (IN-DIACom), 2015, pp. 262-266.

[2] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, Vol. 5, No. 1, pp. 1-167, May 2012.

[3] B. Pang, L. Lee, S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2002, pp. 79–86.

[4] P.D. Turney, "Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews," *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Philadelphia, Pennsylvania, USA, 2002.

[5] B. Wang, M. Liu, *Deep Learning for Aspect Based Sentiment Analysis*, Stanford University report, https://cs224d.stanford.edu/reports/WangBo.pdf, 2015.

[6] M. Hu, B. Liu, "Mining and summarizing customer reviews," *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'2004)*, 2004, pp. 168–177.

[7] M. Hu, B. Liu, "Mining opinion features in customer reviews," *Proceedings of the National Conference on Artificial Intelligence AAAI'04*, San Jose, California, July 25-29, 2004, pp. 755–760.

[8] N. Jindal, B. Liu, "Mining comparative sentences and relations," *Association for the Advancement of Artificial Intelligence AAAI*, Vol. 22, pp. 1331-1336, 2006.

[9] B. Liu, "Sentiment analysis and subjectivity," *Handbook of Natural Language Processing* 2, 627-666, 2010.

[10] K. Khan, B. Baharudin, A. Khan, A. Ullah, "Mining opinion components from unstructured reviews: A review," *Journal of King Saud University – Computer and Information Sciences*, Vol. 26, pp. 258–275, 2014.

[11] R. Studer, V. R. Benjamins, D. Fensel, "Knowledge engineering: principles and methods," *Data & Knowledge Engineering*, Vol. 25, Issue 1, pp. 161-197, 1998.

[12] S. R. Garner, "Weka: The waikato environment for knowledge analysis," *Proceedings of the New Zealand Computer Science Research Students Conference*, 1995, pp. 57-64.

[13] L. Farek, *Identification d'opinions dans les textes arabes en utilisant les ontologies*, PhD Thesis, Université Badji Mokhtar – Annaba, Algeria, 2014. [in French].

[14] L. Zeng, F. Li, "A classification-based approach for implicit feature identification," *Proceedings of the Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, 2013, pp. 190-202.

[15] L. Farek, T. G. Yamina, "Mining explicit and implicit opinions from reviews," *International Journal of Data Mining, Modelling and Management*, Vol. 8, Issue 1, pp. 75-92, 2016.

[16] L. A. Freitas, R. Vieira, "Ontology based feature level opinion mining for Portuguese reviews," *Proceedings of the ACM 22nd International Conference on World Wide Web*, 2013, pp. 367-370.

[17] K. Bafna, D. Toshniwal, "Feature based summarization of customers' reviews of online products," *Procedia Computer Science*, Vol. 22, pp. 142-151, 2013.

[18] B. Bhatt, P. Bhattacharyya, "Domain specific ontology extractor for Indian languages,"

Proceedings of the 10th Workshop on Asian Language Resources, COLING, Mumbai, 2012, pp. 75-84.

[19] V. Narayanan, I. Arora, A. Bhatia, "Fast and accurate sentiment classification using an enhanced Naive Bayes model," Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning, 2013, pp. 194-201.

[20] S.O. Orimaye, S.M. Alhashmi, E-G. Siew, "Performance and trends in recent opinion retrieval techniques," *Knowledge Engineering Review*, Vol. 30, No. 1, pp. 76–105, 2015.

[21] A.K. Samha, Y. Li, J. Zhang, "Aspect-based opinion extraction from customer reviews," arXiv preprint arXiv:1404.1982, 2014.

[22] K. Schoutenm, F. Frasincar, "Finding implicit features in consumer reviews for sentiment analysis," *Proceedings of the International Conference on Web engineering ICWE'2014*, 2014, pp. 130–144.

[23] T. A. Rana, Y. N. Cheah, "Aspect extraction in sentiment analysis: comparative analysis and survey," *Artificial Intelligence Review*, pp. 1-25, 2016.

[24] H. Nakagawa, T. Mori, "A simple but powerful automatic term extraction method," *Proceedings of the International Workshop on Computational Terminology*, Morristown, NJ, USA, 2002.

[25] M. Hu, B. Liu, "Mining opinion features in customer reviews," *Proceedings of the National Conference on Artificial Intelligence AAAI'04*, San Jose, California, July 25-29, 2004, pp. 755–760.

[26] S. Poria, E. Cambria, L. W. Ku, C. Gui, A. Gelbukh, "A rule-based approach to aspect extraction from product reviews," *Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP)*, 2014, pp. 28-37.

[27] S. Mukherjee, J. Ajmera, S. Joshi, "Unsupervised approach for shallow domain ontology construction from corpus," Proceedings of the 23rd ACM International Conference on World Wide Web, 2014, pp. 349-350.

[28] http://sentiwordnet.isti.cnr.it/].

[29] J. Kreutzer, N. Witte, *Opinion Mining Using SentiWordNet*, Uppsala University, 2013.

[30] M. Al-Maimani, N. Salim, A. M. Al-Naamany, "Enhancing opinion mining classification and scoring," Science International, Vol. 27, No. 2, 2015.

[31] S. Bhatia, M. Sharma, K. K. Bhatia, "A novel approach for crawling the opinions from world wide web," *International Journal of Information Retrieval Research (IJIRR)*, Vol. 6, No. 2, pp. 1-23, 2016.

[32] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, P. Kuksa, "Natural language processing (almost) from scratch," *Journal of Machine Learning Research*, No. 12, pp. 2493-2537, 2011.

[33] A. Pak, P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," *LREc*, Vol. 10, pp. 1320-1326, 2010.

[34] V. Umadevi, "Sentiment analysis using Weka," *International Journal of Engineering Trends and Technology (IJETT)*, No. 18, 2014.

Surbhi Bhatia *is currently teaching in K.R. Mangalam University, Gurgaon as an Assistant Professor and has almost 7 years of teaching experience. She is pursuing Ph.D in Computer Science from Banasthali Vidyapith, Rajasthan. She has published 19 research papers in various International/National Journals. Her areas of interests include Data mining, Machine Learning, and Information Retrieval.*

Manisha Sharma *is former Associate Professor of Banasthali Vidyapith, Rajasthan. She has around 18 years of experience. She has done her Ph.D. in Computer Science from Banasthali University and has published more than 30 research papers in reputed national and international journals.*

Komal Kumar Bhatia *received the B.E. and M.Tech. degrees in Computer Science Engineering with Hons. from Maharishi Dayanand University in 2001 and 2004 and Ph. D. in Computer Engineering from YMCA university of Science and Technology. Presently, he is working as Professor and Chairman in Computer science and Engineering department in YMCA Institute of Engineering, Faridabad. His areas of interests are Search Engines, Crawlers and Hidden Web.*

Pragyaditya Das *is currently working as an Engineering Officer at Indian Oil Corporation Limited. He has done his B.tech in Instrumentation and Control Engineering from NIT Trichy. He has published 6 papers in National/International Journals. All his work can be found on https://github.com/jeet1994, his interests consists of Computational Optimization, Coding and Mathematics.*