A Domain Independent Technique to generate Feature Opinion Pairs for Opinion Mining

ASHWINI RAO[1], KETAN SHAH[2]
[1] IT Department, MPSTME
[2] IT Department, MPSTME
NMIMS University
V.L Mehta Road, Vile Parle (W), Mumbai
INDIA
ashwini.rao@nmims.edu, ketan.shah@nmims.edu

Abstract: - The rapid growth of e-commerce and social media services in recent times have opened up many interesting Opinion mining research problems. Among others, Aspect/Feature based Opinion extraction have managed to grab researcher's attention. Towards this direction, many researchers have proposed techniques that are supervised and domain dependent for extracting opinions. In this paper, a novel technique that is unsupervised and domain independent is proposed for generating relevant Feature Opinion pairs with good accuracy. Technique employs grammatical relationships obtained by using typed dependency parsers to further refine the pairs extracted by Part of Speech taggers. It also focusses on words that are verbs and nouns which in some cases imply opinions, unlike other existing work which mainly focusses on a djectives and adverb expressions for Opinion analysis. The proposed technique was tested on 9 data sets of different domains. The result demonstrated a good percentage reduction in number of irrelevant Feature Opinion pairs and the relevancy of retained pairs was found to be considerably high.

Key-Words: - Opinion mining, Supervised learning, Feature Extraction, Dependency parser.

1 Introduction

In this era, along with the rapid growth of web, we see web users evolving with it. This is basically due to the amount of time spent on v arious social networking sites. People now have become more enthusiastic about sharing their ideas, interacting with others, as well as collaborating through various blogs, wikis, online communities etc.

The opportunity to capture this opinionated data and make use of it in various applications starting from Ecommerce to financial market prediction has raised growing interest in the research community. One such task in this area of Sentiment mining is Opinion Summarization, which provides summaries across several dimensions.

The paper focusses on Aspect Level Opinion Summarization, which unlike a simple textual summary of the reviews, deals with generating summaries of specific Aspects/Features of the entity being talked upon. This entire task can be broken down into the following sub tasks.

- Identification of features/aspects from the review document.
- Associating the identified features with opinions.
- Generating summaries

Towards this research direction, the paper proposes a novel technique to extract relevant feature opinion pairs which helps in creating opinion summaries that are concise and more appropriate. Much of the work in this area are supervised techniques and are domain dependent. The proposed technique in the paper is domain independent and unlike other lexical feature based work, it focusses on the contextual information in the review sentences.

The proposed technique was tested on 9 data sets belonging to different domains. Out of these data sets, 5 were from the widely-used corpora for sentiment analysis by Pang et al. [7]. Also 2 other data sets were crawled and remaining were the golden data sets published by Kavita et al. [19] belonging to hotel industry[1].

The paper is organized as follows. Author's contribution is highlighted in Section 2. R elated research in the area of opinion extraction is outlined in Section 3. Section 4 discusses the proposed technique for generating feature opinion pair. Data

---

[1] http://www.kavita-ganesan.com/entity-ranking-data

sets and experimentation results are discussed in Section 5. And Section 6 finally outlines the future plan.

2 Authors Contribution

In the process of building up the feature opinion pair, most of the previous research makes use of a domain dependent, lexicon based approaches. This paper makes its contribution by proposing a novel technique which uses the grammatical structure and dependency relations between words to form the relevant feature opinion pairs. The technique proposed is unsupervised and is independent of the domain. The approach also takes into consideration, words that are tagged as verbs and nouns which in some cases imply opinions.

3 Related Work

Researchers in last few years have taken keen interest in finding the solutions to the challenging problems in Opinion Summarization. Many researchers are working on various aspects in this area.

The task of associating feature with opinions is one of the crucial task in opinion summarization as relevant pairs extracted in this phase influences the summary generated of the reviews. That is, relevant feature opinion pairs would help in generating precise and concise summaries. In this research direction, several techniques for associating opinions with features and generating feature opinion pairs have been proposed by various researchers. Ruihai et al. [4] in their work compared the features with nearby opinion words from the lexicon. If opinions were found, then the distance between the two and their POS pattern was taken into consideration along with a threshold to select only those patterns which frequently occurs. Ana et al. [5] assumed that opinion phrases would be in the vicinity of the features. But however, a vicinity window was not used. When an extracted feature was found in a sentence, the extraction rules were applied to find the potential heads of opinion phrases.

Lun et al. [6] used a dictionary of sentiment words which was iteratively enlarged. The weight of the important sentiment words was computed based on their frequency of occurrence in the document. Opinion polarities were also determined using opinion operators such as say, suggest, present etc. Minging et al. [7] in their work found out that adjectives are usually opinion words. Those opinionated words which are close to the features extracted were used. The orientation of these words was later found out using WordNet dictionary. Sasha et al. [8] used the approach of breaking down the reviews into sentences and phrases. These were determined to be opinionated or not by using both the static and dynamic features. The overall sentiment score of a sentence was then determined based on the individual sentiment score of the opinionated words.

Li Zhuang et al. [9] used a General inquirer (GI) lexicon which also has the semantic orientation of words for finding out initial opinion words. Only the top 100 positive and negative words were used. This list was then enlarged using the synonyms from WordNet. Along with these the opinion words with high frequency were also added to the generated list. Dependency grammar graph was then used to determine the feature opinion pair. Kushal et al. [10] found out in their work that overgeneralization of words degraded the performance of the feature extraction task. The feature with opinion words were extracted using the noun adjective relationship. Based on their frequency they were then grouped together and were made as n grams.

Ahmad et al. [11] made use of rule based approach to determine feature opinion pairs and machine learning for polarity detection. Stanford parser was used for identifying the parts of speech patterns. The feature opinion pair generated had many irrelevant ones which were reduced using a reliability score. An opinion score generator was also used to compute the opinion score of the opinionated words after feasibility analysis. Then some of these pairs were discarded.

Chih et al. [12] used only subjective adjective words from the general inquirer lexicon and prepared a new list. Using this list some of the features which did not occur with the opinions were removed. Auranzeb et al. [13] worked in the direction to reduce the dependency the other approaches had on the manually created domain dependent lexicon. They classified the sentences into subjective and objective ones. And the polarity of subjective sentence was found out using a lexicon dictionary. This polarity was then updated using the sentence structure and contextual feature of each term. It was observed that adjectives and/or adjectives preceded by adverbs were usually opinion words.

Mita et al. [14] in their research stored every feature extracted along with the list of adjectives describing them. Also, the opinion modifiers like not, but etc. were also stored. A SentiWordnet lexicon which had three normalized sentiment scores of positivity, objectivity and negativity was later used for determining the sentiment polarity.

Lipika et al. [15] made use of a knowledge base which contains information about opinion words and their orientation with respect to a domain. They found out that opinion words and modifiers are not restricted to be just adverbs and adjectives. The relationship between the features and opinion words were learnt as linguistic rules through frequent sequential pattern mining. Some opinion phrases which were domain dependent were also captured and stored in the database.

Zhang's et al. [16] work was based on extracting features near the opinion words and extracting opinion words near the features. A dependency parser based relation template was used to identify both simultaneously. A technique of double propagation was used where an initial seed set of opinion words was utilized to find feature which were then expanded.

Ahmad et al. [17] implemented the feature opinion learner module as a rule based system. An information component triplet <f, m, o> consisting of feature, modifier and opinion was found using 6 different dependency rules to tackle different type of sentence structure. HITS algorithm was then applied to these feature opinion pairs to filter out those with low reliability score.

Huayi et al. [18] found out that in some domains, verb expressions imply opinions. They used Open NLP chunker to parse sentence. A verb expression here was the one with sequence of chunks centered at a verb phrase with a noun phrase chunk to its left and an optional adjective, adverb and noun phrase to its right.

4 Proposed technique for generating Feature Opinion pairs

This section describes the proposed approach to extract relevant feature opinion pairs. In opinion summarization, the task of creating relevant feature opinion pair is important as it would be later used for precise summary generation. Among the researchers who are working in this direction, some of them classify sentences into opinionated or not opinionated by using the dictionary of adjectives. These adjectives are then paired with nearby nouns that occur in those sentences. Few others make use of the features extracted in the previous phases and then link it to the nearest adjective. The adjectives are identified either by using the bag of words method or by using the parts of speech tagger. The number of irrelevant feature opinion pairs that get extracted by following the above mentioned approaches is vast. And also, in some cases when the opinions span across sentences, some relevant feature opinion pairs are missed out.

The proposed system design for generating relevant feature opinion pair is shown in Figure 1. The proposed approach for generation/extraction of feature opinion pairs starts with the single and multiword features that were extracted using the automated rule based algorithm [2].

The document is tagged using a Stanford Parts of Speech tagger[2]. As the proposed method is unsupervised and does not depend on any dictionary for finding and validating adjectives, the task of stemming as done by various other researchers in this phase of opinion identification is not carried out. Also, in the proposed technique, since the taggers and parsers are used for opinion extraction, it was found during experimentation that stemming reduces the number of words that are tagged as JJ/JJR/JJS and are adjectives. Experimentation also revealed that adjectives are found to the right of the feature within a nearby distance.

Further, from this tagged document 16 sub documents were created based on the number of features and adjectives in each of the tagged sentence. It was seen that there were certain statements which had no features and no opinionated words. These sentences were then ignored as they were the statements which were not speaking about any feature of a product or any opinion about the feature.

From these sub documents, it was also observed that there were certain statements which either had only features or had only opinions. This gave us an indication that in some sentences the opinions spanned across statements. Therefore, to take care of these features and opinions which may be left out, the task of pronoun resolution was done to a certain extent. From the rest of the documents, feature opinion pairs were extracted, taking into consideration the observation made earlier that adjectives are found mostly to the right of a feature within a nearby distance. Also, in some cases a verb was found separating the feature and the adjective. Conjunction that occur between features and/or adjectives were also taken into consideration when the feature opinion pairs were generated. Example below shows some feature-opinion pairs extracted.

[2] http://nlp.stanford.edu/software/corenlp.shtml

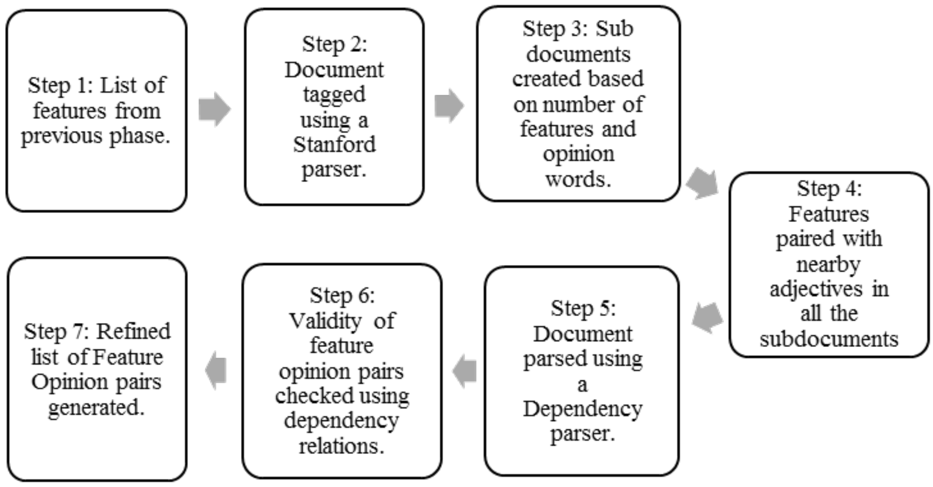

Fig. 1. Proposed system design for generating relevant Feature Opinion pairs

Ex: *The/DT design/NN ,/, as/IN mentioned/VBN above/IN ,/, is/VBZ sleek/JJ ,/, cool/JJ and/CC trendy/JJ -LRB-/-LRB- and/CC I/PRP think/VBP trendy/JJ =/JJ good/JJ ,/, but/CC it/PRP 's/VBZ always/RB hard/JJ to/TO tell/VB these/DT days/NNS -RRB-/-RRB- ./.*

Feature Opinion pairs extracted - *(design,sleek) (design,cool) (design,trendy) (design,hard)*

In the above example, DT, NN, IN, VBZ etc. are the POS tags i.e. the lexical category to which the word in the sentence belongs to.
Many number of such feature opinion pairs obtained were found to be irrelevant. As in the above example, the pair *(design, hard)* is irrelevant. It was mainly due to the task of just combining the features with nearby adjectives irrespective of whether the opinion is truly meant for that feature or not. Therefore, to validate these pairs and retain only the relevant ones, the document was parsed using a Dependency parser[3]. Out of a set of approximately 50 grammatical relations only a few such as nsubj, amod, advmod etc. which can be used to identify features and/or adjectives were considered. Each of these dependencies are binary relations between a dependent and a governor also called as a head [26]. Various associations that occur between these relations were used to check whether the opinions were specific to the features in the feature opinion pair generated previously.

[3] http://nlp.stanford.edu/software/lex-parser.shtml

In some cases, it was found that features were the subject or the object of a nearby verb. This indicates that opinions in these sentences were expressed using verbs. Some examples of the parser relation patterns derived to generate feature opinion pairs is shown in the table below.

Table 1. Examples of Dependency Relation Patterns

Parser Dependency	Feature word	Opinion word
amod(quality-2, sound-1);nsubj(great-4, quality-2); cop(great-4, is-3);	sound quality	great
nsubj(comes-4, it-3); amod(headset-10, sleek-7); conj:and(sleek-7, powerful-9); amod(headset-10, powerful-9)	headset	sleek, powerful
det(service-2, The-1);nsubj(good-8, service-2); amod(the-4, mobile-5); cop(good-8, is-6); advmod(good-8, very-7);	service	very good

The use of dependency rules helps in identifying the scope and context of adjectives. Based on these rules, many irrelevant pairs were eliminated. The example below shows the feature opinion pairs extracted from the same review statement discussed earlier, after applying the proposed technique on the document parsed by a dependency parser.

Ex: [det(design-2, The-1), nsubj(sleek-9, design-2), nsubj(cool-11, design-2), nsubj(trendy-13, design-2), mark(mentioned-5, as-4), dep(design-2, mentioned-5), nmod(mentioned-5, above-6), cop(sleek-9, is-8), root(ROOT-0, sleek-9), conj:and(sleek-9, cool-11), cc(sleek-9, and-12), conj:and(sleek-9, trendy-13), cc(think-17, and-15), nsubj(think-17, I-16), dep(sleek-9, think-17), xcomp(think-17, trendy-18), dep(good-20, =-19), ccomp(think-17, good-20), cc(think-17, but-22), nsubj(hard-26, it-23), nsubj(tell-28, it-23), cop(hard-26, 's-24), advmod(hard-26, always-25), dep(sleek-9, hard-26), conj:and(think-17, hard-26), mark(tell-28, to-27), xcomp(hard-26, tell-28), det(days-30, these-29), nmod:tmod(tell-28, days-30)]

Feature Opinion pairs extracted - *(design,sleek) (design,cool) (design,trendy)*

The pair *(design, hard)* that was extracted earlier is marked as irrelevant by the proposed technique as the adjective *hard* is not being used in the context of any product feature. Likewise many such irrelevant pairs are eliminated to get a refined and concrete list of feature opinion pairs which would help us to generate relevant summaries.

The evaluation of the proposed technique of feature-opinion pair generation on the different data sets is discussed in the next section.

5 Experiments

5.1 Corpus

The experiments were conducted on 9 different data sets. Of these, five were customer reviews for products such a Nokia 6600, Norton Antivirus, Router, IPOD, and Micro MP3. These were from the standard corpus used by Pang et al. [7] for sentiment analysis. Each of these reviews have around 8 to 9 sentences and an average of 7 tokens per sentence. Also, 2 other data sets were from the domain of Hotel industry[1] and published by Kavita et al. [19]. Compared with the earlier data sets, the reviews here are considerably shorter. Each review had around 2 to 3 sentences and an average of 13 tokens per sentence. The features and their associated opinion strengths were annotated in the above mentioned golden data sets by the publishers. A sample of the annotated data set is shown below.

Sample of Annotated Data Set:

ringtone[+1],background[+1], screensaver[+1], memory[-2]## the phone comes with okay ringtones , some decent backgrounds / screensavers , but the phone has very little memory (mine had 230kb as it arrived from amazon , so you do n't have too many options on what you can put on there) .

Various tags as shown above are used in these annotated reviews. Such as, *xxx [+/-n]* indicating the opinion strength n of a particular feature *xxx* and *##* depicting the start of each sentence.

Apart from the data sets described above which are the golden data sets that are published by respective researchers, the proposed technique was also tested on two more data sets that were crawled from web sites such as CNET, Amazon, Team-bhp and Carswale.com. The reviews were from the domain of Automobile industry and that of Mobile phones. Around 450 reviews were crawled with an average of 15 sentences per review and 7 tokens per sentence.

In the above discussed golden data sets, only the features and their orientation were manually annotated by the research publishers. So as to justify the feature opinion pairs extracted using the proposed approach, all the 9 data sets had to be reviewed and feature/s with their corresponding opinion/s were tagged manually.

5.2 Results

The experimentation starts with the set of single and multiword features that were extracted using the proposed automated rule based approach [3]. As discussed in section 4 earlier, the review document is first tagged using a Stanford POS Tagger[2]. Also, stemming was not done as it reduced the number of adjectives found. Experimentation results of the percentage reduction in number of adjectives after Stemming is shown in Table 2.

Table 2. Reduction in percentage of Adjectives found

Data Sets	% of missed adjectives
iPod	6%
Router	4%
MP3	9%
Norton	3%
Nokia_6610	11%
Cars	15%
Hotel 1	20%
Hotel2	17%
Nokia Lumia	4%

The document is now divided into 16 subdocuments, based on the number of features and adjectives in every sentence as shown in Table 3 below.

Table 3. Created Subdocuments

0 feature – 0 adjective	1 feature – 0 adjective
0 feature – 1 adjective	1 feature – 1 adjective
0 feature – 2 adjectives	1 feature – 2 adjectives
0 feature – Many adjectives	1 feature – Many adjectives
2 features – 0 adjective	Many features – 0 adjective
2 features – 1 adjective	Many features – 1 adjective
2 features – 2 adjectives	Many features – 2 adjectives
2 features– Many adjectives	Many features – Many adjectives

Out of these documents, the ones with 0 features and 1 t o many adjectives and the ones with 0 adjectives and 1 to many features were handled by pronoun resolution. The document with 0 feature and 0 adjective was ignored, as the sentences here are the ones which have no features and are not opinionated.

This stage 1, starts with these sub documents, and the features list from the previous phase [3]. The words that are tagged as ad jectives (i.e. tagged as JJ/JJR/JJS) and which occur to the right of these features within a d istance of 10 o r less than 10 words were extracted to form the feature adjective pairs. Along with these adjectives, some verbs, nouns and adverbs which occur in a specific pattern along with the features were also considered to be the opinion words. The rules involving the verbs, nouns and adverbs for extracting feature opinion pair is as shown in Table 4 below.

Table 4. Relation pattern used to form Feature Opinion Pairs

Relation Pattern	Feature	Opinion
NN - NNS – RB	NN	NNS
NNP -VBZ – RB	NNP	VBZ, RB
NNS – JJ – VBN	NNS	JJ,VBN
RB – VBZ – JJ	RB	VBZ
NN – VBZ - NNP	NN	NNP

The total number of feature opinion pairs generated across various data sets at this stage 1, is as shown in the Table 5 below.

Table 5. Number of Feature Opinion Pairs generated in stage 1

Data Sets	Total Number of Feature Opinion Pairs
iPod	653
Router	614
MP3	1182
Norton	345
Nokia 6610	853
Cars	3210
Hotel 1	2686
Hotel 2	3734
Nokia Lumia	3889

It was observed that, this technique may also have generated feature opinion pairs that are irrelevant ones. This was mainly due to the fact that feature opinion pairs were formed without taking into consideration the context in which the opinion words were used.

In order to filter out such irrelevant pairs generated, the documents were now parsed using a Stanford Parser[3]. Dependency relations as discussed in the previous section were used. Each of these, are binary relations between a head and a dependent. So, out of a set of around 50 grammatical relations, only a few DP relations such as nsubj, amod, advmod, det, cop etc. which had the feature as the dependent or as a head in these binary relations were used to validate the feature opinion pairs extracted previously. That is, only those pairs in which the opinion words were truly meant for the feature in the pair were retained.

The Table 6 shows the number of pairs that were classified as r elevant and irrelevant from the set of pairs that were generated previously using the proposed technique. As can be seen in the Table 6, the number of irrelevant pairs in the last four data sets is considerably high. This is because, the data set is vast and the number of tokens that make up a sentence is also high. These 4 are the data sets that were crawled from various blogs and customer reviews from many popular ecommerce web sites.

Table 6. Number of Relevant & Irrelevant pairs generated using the proposed technique

Data Sets	Relevant Pairs	Irrelevant Pairs	Total Pairs
iPod	141	512	653
Router	139	475	614
MP3	289	893	1182
Norton	67	278	345
Nokia_ 6610	207	646	853
Cars	326	2884	3210
Hotel 1	428	2258	2686
Hotel 2	611	3123	3734
Nokia Lumia	633	3256	3889

The pairs so generated were then checked against the pairs that were manually extracted. The Table 7 below shows the percentage of pairs that were matching with the manually tagged ones.

Table 7. % of pairs matching with manually tagged ones (Proposed technique)

Data Sets	% of matching pairs
IPOD	76%
Router	85%
MP3	86%
Norton	78%
Nokia_6610	90%
Cars	92%
Hotel 1	91%
Hotel 2	90%
Nokia Lumia	94%

It was observed that the proposed technique also extracted some pairs which were not a p art of manually extracted ones. This for some extent was due to fact that wrong dependency relations were generated because of the error in the syntactic structure of the sentence. The example below depicts a syntactically wrong sentence and the binary relations generated using a dependency parser.

Ex: IPod is brilliant, but service was bed Wow iPod.

After being parsed:
[nsubj(brilliant-3, IPod-1), cop(brilliant-3, is-2), root(ROOT-0, brilliant-3), cc(brilliant-3, but-5), nsubj(iPod-10, service-6), cop(iPod-10, was-7), compound(iPod-10, bed-8), compound(iPod-10, Wow-9), conj:but(brilliant-3, iPod-10)]

The example above shows that the word *bed* used in the above context is syntactically wrong and is tagged with the binary relation *compound*. The correct word should have been *bad*, and accordingly it would have been tagged with the *amod* relation.
The percentage of such irrelevant pairs in data sets that were unstructured was found to be relatively high. This could have been avoided, if some additional spell correction had been carried out based on t he context in which the word is being used in a sentence.
A lexicon based approach of pairing features with nearby adjectives was implemented. In order to check the relevancy of the pairs extracted using the lexicon approach, these feature opinion pairs were then compared with the manually tagged ones. The results of the number of pairs extracted and the percentage of pairs matching with the manually tagged ones is tabulated in Table 8.

Table 8. Number of Feature Opinion pairs extracted using the Lexicon/Dictionary approach

Data Sets	Total Number of Feature Opinion Pairs	% of matching pairs
IPOD	888	50%
Router	734	56%
MP3	1487	54%
Norton	447	54%
Nokia_6610	1218	64%
Cars	5377	67%
Hotel 1	4476	68%
Hotel 2	5939	64%
Nokia Lumia	6284	67%

The result clearly highlights that the number of mismatched pairs is relatively high when compared with the proposed technique and that of pairs matching with the manually tagged ones is considerably low.

6 Conclusion & Future Work

As the various techniques used for opinion extraction are domain dependent and supervised methods, the paper proposes a novel domain independent approach to generate feature opinion pairs. The technique proposed takes Parts of Speech

as well as dependency relations between words of a sentence into consideration when feature opinion pairs are generated. In order to handle cases in which opinions span across sentences, pronoun resolution was also done. Along with this, during the task of opinion word extraction, the verbs and nouns were also examined. This was mainly done to handle sentences that had words tagged as v erbs and/or nouns and were opinionated.
The proposed technique was evaluated on varied data sets belonging to different domains. The results demonstrated a good number of relevant feature opinion pairs that were extracted. This would help us in generating the opinion summaries of these reviews which would be concise as well as precise.

References:

[1] B. L. Minqing Hu.: Mining Opinion Features in Customer Reviews. *American Association for Artificial Intelligence (AAAI*), pp. 755–760. (2004)

[2] Rao, Ashwini, and Ketan Shah. : Model for Improving Relevant Feature Extraction for Opinion Summarization. In *IEEE International Conference on Advance Computing (IACC)*, pp. 1-5. IEEE. (2015)

[3] Rao, Ashwini, and Ketan Shah.: An optimized rule based approach to extract relevant features for sentiment mining. In *Proceedings of 3rd International Conference on Computing for Sustainable Global Development (INDIACom)* IEEE. (2016)

[4] Dong, R., Schaal, M., O'Mahony, M. P., & Smyth, B.: Topic extraction from online reviews for classification and recommendation. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pp. 13 10-1316, AAAI Press. (2008)

[5] Popescu, A. M., & Etzioni, O.: Extracting product features and opi nions from reviews. In *Natural language processing and text mining (pp. 9-28)* Springer London. (2007)

[6] Ku, L. W., Liang, Y. T., & Chen, H. H.: Opinion Extraction, S ummarization and Tracking in News and Blog Corpora. In *AAAI spring symposium: Computational approaches to analyzing weblogs*, Vol. 100107. (2006)

[7] B. Pang and L. Lee.: A sentimental education: Sentiment analysis using subjectivity summarization based on minimum Cuts. In *Proceedings of the ACL*, pages 271–278. ACL.(2004)

[8] K. H. Sasha Blair-Goldensohn, Ryan McDonald,Tyler Neylon,George A. Reis,Jeff Reynar. : Building a S entiment Summarizer for Local Service Reviews. In *International World Wide Web Conference Committee (IW3C2),* Beijing, China.(2008)

[9] Zhuang, L., Jing, F., & Zhu, X. Y.: Movie review mining and summarization. In *Proceedings of the 15th ACM international conference on Information and knowledge management* (pp. 43-50), ACM. (2006).

[10] Dave, K., Lawrence, S., & Pennock, D. M.: Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web* (pp. 519-528), ACM. (2004)

[11] A Abbasi, HC Chen and A Salem.: Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems*, Vol. 26, No. 3. (2008)

[12] C.-P. Wei, Y.-M. Chen, C.-S. Yang, and C. C. Yang. : Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews. *Information Systems and e-Business Management*, vol. 8, pp. 149-167. (2009)

[13] Khan, Aurangzeb, Baharum Baharudin, and Khairullah Khan. : Sentiment classification from online customer reviews using lexical contextual sentence structure. *International Conference on Software Engineering and Computer Systems*. Springer, Berlin, Heidelberg. (2011)

[14] M. K. Dalal and M. A. Zaveri.: Opinion Mining from Online User Reviews Using Fuzzy Linguistic Hedges. *Applied Computational Intelligence and Soft Computing*, Vol. 2014, pp. 1-9. (2014)

[15] L. Dey and S. M. Haque.: Opinion mining from noisy text data. *International Journal on Document Analysis and Recognition (IJDAR)*, Vol. 12, pp. 205-226. (2009)

[16] T. Y. Wen Zhang. Xijin Tang. : Text Classification using Multi-word Features. *IEEE*, pp. 3519-3524. (2007)

[17] M. A. Z. Mita K. Dalal. : Semisupervised Learning Based Opinion Summarization and Classification for Online Product Reviews. *Applied Computational Intelligence and Soft Computing*, p. 8. (2013)

[18] Li, Huayi, Arjun Mukherjee, Jianfeng Si, and Bing Liu. : Extracting Verb

Expressions Implying Negative Opinions. In *AAAI*, pp. 2411-2417. (2015)
[19] Ganesan, Kavita, and Chengxiang Zhai. : Opinion-based entity ranking. *Information retrieval* 15.2 pp. 116-150. (2012)
[20] P. S. Yelena Mejova, Bob Boynton. : GOP Primary Season on T witter: “Popular” Political Sentiment in Social Media. I n *WSDM*, pp. 517-526. (2013)
[21] Yang, Y., & Pedersen, J.O.: A comparative study of feature selection in text categorization. In *Proceedings ICML*, pp. 412-420. (1997)
[22] R. T. N. Giuseppe Carenini, Ed Zwart. : Extracting Knowledge from Evaluative Text. I n *KCAP'05*, Banff, Alberta, Canada, p. 1595931635/05/0010. (2005)
[23] Kobayashi, Nozomi. : Collecting evaluative expressions for opinion extraction. *Natural Language Processing–IJCNLP 2004*. Springer Berlin Heidelberg, pp. 596-605. (2004)
[24] Carenini, Giuseppe, Raymond T. Ng, and Adam Pauls.: Multi-Document Summarization of Evaluative Text. *EACL*. (2006)
[25] T. Y. Wen Zhang, Xijin Tang. : Text Classification using Multi-word Features. *IEEE*, pp. 3519-3524. (2007)
[26] De Marneffe, Marie-Catherine, and Christopher D., Manning. : *Stanford typed dependencies manual. Technical report*, Stanford University. (2008)
[27] Ashwini Rao, Dr. Ketan Shah. : Filtering and Transformation Model for Opinion Summarization. *International Journal of Computers & Technology*, Vol. 13, No. 2, pp.4248-4255. (2014)