© 2016 IJSRSET | Volume 2 | Issue 3 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Fine Grained Opinion Mining from Online Reviews for Product Recommendation

R. S. Gaikwad, Dr. B. L. Gunjal

Department of Computer Engineering, Amrutvahini College of Engineering, Sangamner, Maharashtra, India

ABSTRACT

Fine grained opinion mining is an important task in today's E-business world. Customers and manufacturers wanted to know about products in details. So in this paper we have studied opinion targets and opinion word extraction through dependency parsing and by applying syntactic patterns. Previously developed double propagation approach is useful for extraction task with addition of some syntactic relations. Also finding opinion orientation can be performed using dictionary and some contrary words, conjunctions. We can also generate a summary to analysis and recommendations. By using opinion target list and opinion word lexicon we can achieve better results.

Keywords: Opinion target, Opinion words, Opinion orientation, Recommendation.

I. INTRODUCTION

With the fast growth of technology and internet services online shopping is increasing rapidly. Customers are keen to purchase a product with good quality and features. But before buying a product customers always used to check the reviews of products and its features in which they are interested in. Customers always make decisions based on previous reviews about a product. Not only customers but manufacturers also interested in finding reviews about their product to know the quality and defects in their products and make improvements in their products. So mining online reviews is increasingly important task for today's customers and also for manufacturers. To analyse and extract opinion from online reviews it is important to obtain fine-grained sentiment about features and aspect from online reviews. For Example:

"I like this delicious food, but disgusting service"

In the above example the customer express negative opinion about services and positive opinion about food. So other customers must come to know fine-grained opinion about things. So in this paper obtaining fine-grained features of product and its associated opinion is major task. So we extract opinion targets and opinion words from online reviews.

Users express their opinion about its feature or attribute is called as opinion target. In above example "food" and "service" are opinion targets. Opinion targets or features/attributes are generally noun or noun phrases. This task is also called as product feature extraction. Opinion words are words which are used to express opinion about feature. In above example "delicious" and "disgusting" are opinion words.

To extract opinion targets and words previous methods used word alignment models, Collective extraction strategy, syntactic patterns, nearest neighbour. In collective extraction method it is observed that opinion words and opinion targets co-occur with each other. So there must be a strong modification relation and opinion association among them. This can be called as opinion relation [1]. So some collective strategies extracted them in bootstrapping way. In above example "delicious" used to modify "food" and "disgusting" is used to modify "service".

When we get opinion targets and opinion words and their association, we have to find out opinion orientation of that feature. Taking same example "delicious food" is

positive opinion for the food and "disgusting service" is negative opinion. So our task is to find out whether customers have expressed positive or negative opinion about a feature. Sentiment is associated with the sentiment word in opinion word lexicon. We have to determine some addition things such as "not good" is negative opinion.

Opinion summary generation is also important task in opinion mining. How many number of positive or negative opinion associated with the features. It will be helpful to the customers and also to manufacturer of product. Summary may be of different types based on the application.

And finally all these can be useful to recommend a product to a customer. If any one customer interested in specific feature he/she can enter a keyword or a sentence for that and based on that we can recommend result from our analysis.

So in this paper we have studied some major tasks to be performed.

(1) Efficiently extract opinion targets and opinion words and their association.
(2) Identify opinion orientation about a feature, whether it is +ve or –ve.
(3) Opinion summary generation about all extracted features.
(4) And finally recommend a product with required feature.

II. METHODS AND MATERIAL

1. Related Work

K. Liu, L. Xu, and J. Zhao [1], in this paper an alignment-based approach with graph co-ranking to collectively extract opinion targets and opinion words is proposed. To detect opinion relations among words they used syntactic parsing and a monolingual word alignment model (WAM) with EM-based hill climbing algorithm. An opinion target can find its corresponding modifier through word alignment. Then they used graph co-ranking to estimate candidate confidence. Domain-independent general noun (GN) corpus used to filter out noise. For experimental purpose they used Minipar parser for English reviews and Stanford parser for Chinese reviews, SentiWordNet as external manually labeled resource, POS-tagger and three datasets were used.

M. Hu and B. Liu [2], In this paper they focused on mining product features, identifying opinion sentences, finding opinion orientation and summarizing the result. For finding frequent features association mining base in Apriori algorithm used. They used POS-Tagging to know part-of-speech of words. They find nouns and adjectives as near distance between them, Wordnet for finding out opinion orientation of words. A count is computed to show how many reviews give positive/negative opinions to the feature. A count is computed to show how many reviews give positive/negative opinions to the feature.

F. Li, S. J. Pan, O. Jin, Q. Yang, X. Zhu [3], proposed a collective extraction method for extracting opinion targets and words. They used patterns and from source domain to extract topic and words of target domain using Relational Adaptive bootstrapping (RAP) algorithm. It does not require labelled data.

L. Zhang, B. Liu, S. H. Lim, E. O. Strain [4], they focused on mining features. To increase the recall they introduced part-whole and "no" patterns. Feature relevance and feature frequency used to know importance of feature. The system uses predefined part-whole relation patterns to extract features in a domain corpus. For experiment Cars, Mattress, Phone, LCD datasets they used.

K. Liu, L. Xu, J. Zhao [5], proposed system to find opinion relations between nouns/noun phrases and adjectives in sentences, and calculate the associations between them. For this task they used a IBM-3 word alignment model and graph algorithm. For experiments they used POS-tagging and CRD dataset.

M. Hu and B. Liu [6], proposed a system to summarize the customer reviews of a product. Also opinion orientation identification the opinions are positive or negative using a bootstrapping technique and the WordNet. They used Part-of-Speech Tagging, Feature Pruning methods. For frequent feature they extract the near *adjectives.* Dataset of 2 digital cameras, 1 DVD player, 1 mp3 player, and 1 cellular phone were used for experiments.

A. Popescu and O. Etzioni [7], proposed OPINE, an unsupervised information extraction system. This system mines reviews to build a model of important product features. It uses a *relaxation labelling* for finding the semantic orientation of words. Also based on their strength they Ranked opinions. They used MINIPAR parser. For extraction they used relation-specific generic extraction patterns into extraction rules.

G. Qi, B. Liu, J. Bu, C, Chen [8], proposed system for opinion lexicon expansion and opinion target extraction. It requires only an initial seed opinion lexicon hence also called as weakly semi-supervised method. There is an opinion relations between opinion words and targets because opinion words are used to modify targets. Opinion targets and words extract iteratively using known opinion targets and words. They used Minipar as a dependency parser for the identification of syntactic relations. Opinion word polarity detection and noisy target pruning methods are also used to correct and refine the initially extracted results. This approach propagates information back and forth between words and target, Hence they call it double propagation.

B Wang, H Wang, [9] they considered the problem of identifying product features with opinion words and learning opinion words through features iteratively and alternately. Words and targets are having association between them and a revised formula for mutual information is given in this paper. They have also given a mapping function from words to features to identify implicit features in sentence. Dataset of digital camera, cell-phone and tablet were used for experimental purpose.

G. Qi, B. Liu, J. Bu, C, Chen [11], In this paper, they proposed a propagation method that gives the relations between sentiment words and product features. There is modification relation between sentiment word and a future. So this method can extract sentiment words using sentiment words; extract features using features; extract sentiment words using features; extract features using sentiment words. They used dependency parser to know the dependency between words and generate a dependency tree. Based on relations described in dependency trees the extraction rules were designed. They used POS-Tagging using Stanford POS-Tagger, Minipar as a dependency parser, syntactic relations and double propagation algorithm to extract features and opinion/sentiment words.

X. Ding, B. Liu, P. S. Yu [13], studied the problem of finding opinion orientation opinion words on features in product reviews. They proposed a holistic lexicon-based approach to solving the problem by exploiting external evidences and linguistic conventions of natural language expressions. This method handles opinion words that are context dependant. Also handles phrases, language constructs and special words.

F. Li, C. Han, M. Huang, X. Zhu, Y. Xia, S. Zhang and H. Yu [14], proposed a method Conditional random field (CRF) based machine learning framework. Their method jointly extracted opinion words and targets. They also studied syntactic tree structure and conjunction structure. They used a product review dataset and a movie review dataset for experimental purpose. Totally 601 reviews from four products and 500 reviews from five movies were used.

Y. Wu, Q. Zhang, X. Huang, L. Wu [15], proposed method for opinion mining for identifying opinion targets, opinion expressions and relation between them. They used a phrase dependency parsing. Generally there are some product features and the names of products are composed of more than one word which is noun phrase. So accuracy of parsing can be increased. Dataset of customer reviews of 11 products from 5 categories Diaper, Cell Phone, Digital Camera, DVD Player, and MP3 Player taken for experiment.

2. Overview of System

Our first task is to extract opinion targets and opinion words. Generally there is a modification relation between opinion targets and words. Opinion targets are generally noun/noun phrases and opinion words are adjectives/verbs [1], [5], [8], [11]. We have to identify opinion association between them. We apply opinion orientation logic for extracted targets and words to find out whether it is +ve or –ve. We are checking for some contrary words “not”, “but”, “however” to find opinion orientation [2], [7], [18]. Then we are generating the summary of all the orientation results about a feature [2], [6]. Then apply finally recommendation logic [19].

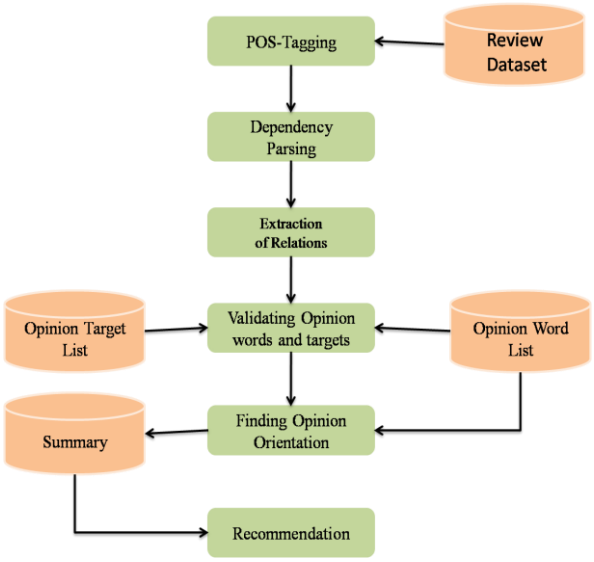

Figure 1 : Architecture.

3. Opinion words and Target extraction

We know that opinion words are co-occur with opinion targets and there is strong modification relation between them [1]. Means opinion word tells more about opinion targets how they are. Opinion words which are adjective/verbs and opinion targets are noun /noun phrases.

A. POS-Tagging

So we have to fist find out part-of-speech of each word in a sentence. For this task we fist employ a POS tagging to our opinion sentences, which gives us part-of-speech of each word. We are using Stanford POS-Tagger in our system. From the observation of POS output we come to know that words in noun phrases are separately tagged as noun. Ex: "battery life" is separately tagged as nouns. But it is a single target. Many feature compaction methods are previously used [6], [8].

B. Dependency Parsing

To find out dependencies between words we can sort our sentences to dependency parsing. For this task previous methods used dependency parser [1], [5], [8], [11]. There may be problem with parser because there may be grammatical mistakes and infrequent writings in reviews. But still we can obtain some useful relations through parser [1]. We consider only direct dependency relations for our task.

In above fig2 (a) A is directly dependant on B. in figure (b) A and B they are directly dependant on the third word C. These are direct dependency relations. In figure (c) A is indirectly dependant on B through additional word C and in figure (d) A and B both Dependant on D through words C1 and C2. These are indirect dependencies. As we have mentioned earlier that we are only considering direct dependencies in our task. We used Stanford dependency parser for this task.

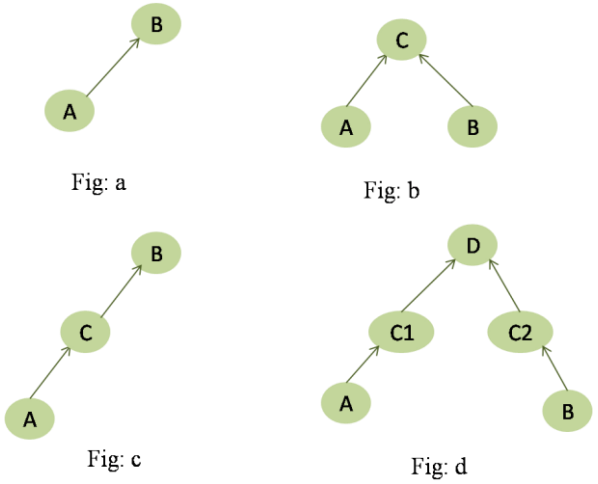

Figure 2: Direct and In-Direct Relations

C. Applying Syntactic Patterns

We used direct dependency relations {amod; rcmod; nsubj; nsubjpass;} between noun/noun phases and adjective/verb. As mentioned earlier that we get large number of pairs of opinion targets and opinion words through syntactic relations or syntactic patterns. We confirm our targets by using opinion target list which we have generated manually. There may be some false targets in our list so they can be automatically removed with the verification of previous list. People generally compare products and features with other produces. So other products are not our targets and we have to eliminate them.

4. Opinion Orientation Identification

Now we have got our opinion targets and their associated opinion words. Our next task is to find opinion orientation of these targets. Opinion words have their polarity. So their opinion targets can get the same polarity as opinion words. But we have to check some other words around them. Some examples of such words are "not", "n't", "despite", "though", "however", "but", "except", "oddly", "aside" and "although" negation

words [8]. These words can change the opinion polarity of opinion words. We can find orientation of opinion words from positive negative word dictionary. We have to check negation and conjunction words in addition. By performing simple calculation with assigning 1 to positive words and -1 for negative words we get the value as positive or negative for the target. If we got positive value then target orientation is positive otherwise negative.

5. Summary Generation

Opinion summary generation is nothing but calculating number of +ve and –ve opinion about a feature in reviews [2], [6]. We have maintained the list of all targets, number of positive opinions, number of negative opinions separately. When we detect the positive polarity of the target we increase the count in positive column in front of target. When we detect the negative polarity of the target we increase the count in negative column in front of target. Then we get the total number of positive and negative count about the target. This will be useful in recommending the product to the customer.

6. Recommendation

Now our final task is to recommend a product to a customer. As our system is web based user can search a product through a query. A query may be in the form of keywords or sentences like reviews. We extract targets from the user query and find the product in our database of summary. We recommend a product to a customer which is having good reviews for the keywords also and can consider the summary of other aspects of the product. We are not going to discuss in details here [refer 19]. But the only thing is our analysed result can be very useful in recommending products.

III. RESULTS AND DISCUSSION

1. Experimental Setup

To extract opinion targets and words we first perform POS-Tagging using Stanford POS-Tagger. So we can get parts-of-speech of each word in review. Then to perform dependency parsing Stanford parser used. Then by developing syntactic patterns we can extract opinion targets and words. We can use previously manually generated opinion target list to verify correct opinion targets and to eliminate false if any. Then to know opinion orientation of words we can use positive negative word dictionary. We used a review dataset of five products in experiments.

2. Results

We can achieve better results using syntactic patterns for opinion targets and words extraction. Also use of opinion target list which is manually generated from observation and knowledge is useful for verification and ambiguity removing purpose. False targets can get eliminated. We can achieve results as shown in table i.e. Precision, Recall and F-measure.

TABLE I : RESULTS

Dataset	Precision	Recall	F-measure
Data1	0.88	0.82	0.85
Data2	0.91	0.83	0.87
Data3	0.90	0.85	0.87
Data4	0.83	0.82	0.82
Data5	0.92	0.91	0.90

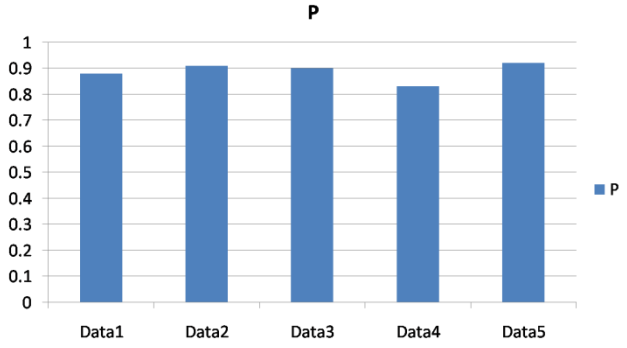

Fig. 3. Precision

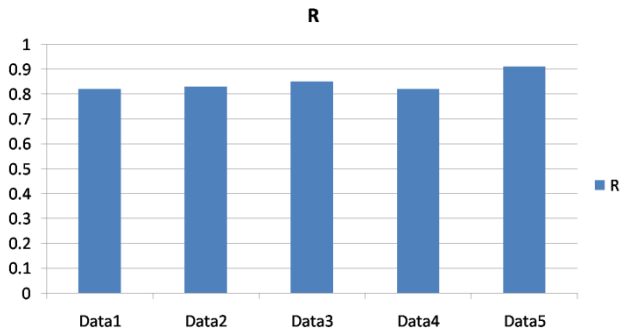

Fig. 4. Recall

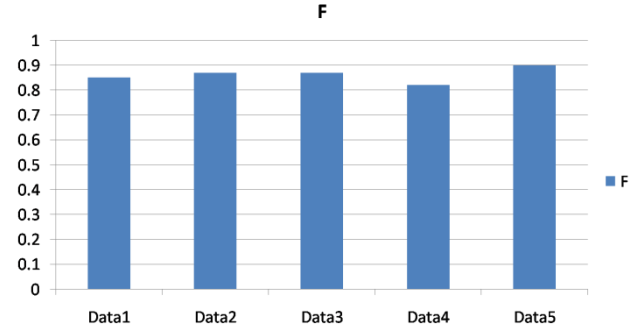

Fig. 5 F-measure

IV. CONCLUSION

Through this study we conclude that opinion words and opinion targets extraction is important task for fine grained opinion mining. Customers and manufacturers can get appropriate information about product features. Use of dependency parsing is important to extract opinion targets and opinion words. Then finding out opinion orientation is also an important task to generate summary. By using target list and sentiment word dictionary we can get true opinion targets and words and can improve results. And finally we can recommend products to the customer through our summary.

V. ACKNOWLEDGMENT

I am very much thankful of Dr. B.L. Gunjal for their guidance and consistent encouragement in this paper work.

VI. REFERENCES

[1]. S. M. K. Liu, L. Xu, and J. Zhao, "Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model" in IEEE Transactions on Knowledge and Data Engineering, Vol. 27, no. 3, pp.636-650, March 2015.

[2]. M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Seattle, WA, USA, pp. 168–177, 2004.

[3]. F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu, "Cross-domain co-extraction of sentiment and topic lexicons," in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Jeju, Korea, pp. 410-419, 2012.

[4]. L. Zhang, B. Liu, S. H. Lim, and E. O'Brien-Strain, "Extracting and ranking product features in opinion documents," in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, pp. 1462–1470, 2010.

[5]. K. Liu, L. Xu, and J. Zhao, "Opinion target extraction using word based translation model," in Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn., Jeju, Korea, pp. 1346–1356, Jul. 2012.

[6]. M. Hu and B. Liu, "Mining opinion features in customer reviews," in Proc. 19th Nat. Conf. Artif. Intell., San Jose, CA, USA, pp. 755–760, 2004.

[7]. A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process., Vancouver, BC, Canada, pp. 339–346, 2005.

[8]. G. Qiu, L. Bing, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," Comput. Linguistics, vol. 37, no. 1, pp. 9–27, 2011.

[9]. B. Wang and H. Wang, "Bootstrapping both product features and opinion words from chinese customer reviews with crossinducing," in Proc. 3rd Int. Joint Conf. Natural Lang. Process., Hyderabad, India, pp. 289–295, 2008.

[10]. B. Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, series Data-Centric Systems and Applications. New York, NY, USA: Springer, 2007.

[11]. G. Qiu, B. Liu, J. Bu, and C. Che, "Expanding domain sentiment lexicon through double propagation," in Proc. 21st Int. Jont Conf. Artif. Intell., Pasadena, CA, USA, pp. 1199–1204, 2009.

[12]. R. C. Moore, "A discriminative framework for bilingual word alignment," in Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process., Vancouver, BC, Canada, pp. 81–88, 2005.

[13]. X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in Proc. Conf. Web Search Web Data Mining, pp. 231–240, 2008.

[14]. F. Li, C. Han, M. Huang, X. Zhu, Y. Xia, S. Zhang, and H. Yu, "Structure-aware review mining and summarization." in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, pp. 653–661, 2010.

[15]. Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," in Proc. Conf. Empirical Methods Natural Lang. Process., Singapore, pp. 1533–1541, 2009.

[16]. T. Ma and X. Wan, "Opinion target extraction in chinese news comments." in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, pp. 782–790, 2010.

[17]. Q. Zhang, Y. Wu, T. Li, M. Ogihara, J. Johnson, and X. Huang, "Mining product reviews based on shallow dependency parsing," in Proc. 32nd Int.

ACM SIGIR Conf. Res. Develop. Inf. Retrieval, Boston, MA, USA, pp. 726–727, 2009.

[18]. W. Jin and H. H. Huang, "A novel lexicalized HMM-based learning framework for web opinion mining," in Proc. Int. Conf. Mach. Learn., Montreal, QC, Canada, pp. 465–472, 2009.

[19]. S. Meng, W. Dou, X. Zhang, J. Chen, "KASR: A Keyword-Aware Service Recommendation Method on MapReduce for Big Data Applications" IEEE Transactions On Parallel And Distributed Systems, Tpds-2013-12-1141, pp. 1-11, 2013.

BIOGRAPHY

Mr. R. S. Gaikwad is Pursuing Master in Engineering from Amrutvahini College of Engineering Sangamner. Received BE degree from Savitribai Phule Pune university. His interest areas are Data Mining, Big Data, Software Engineering.

Dr. B. L. Gunjal is Associate Professor in computer Engineering and Head of Information Technology Department in Amrutvahini College of Engineering, Sangamner. Received PhD degree in Computer Engineering from Savitribai Phule Pune university. Also received **"BEST Teacher Award"** by Savitribai Phule Pune university. Her Research interest in Image Processing, Data Mining.