OPINION TARGETS AND OPINION WORDS EXTRACTION USING MODIFIED WORD ALIGNMENT MODEL

Vibhuti I. Patel[1], Prof. Mital Panchal[2]

[1] *P.G. student, Department of Information Technology, L.D. College of Engineering, Gujarat, India*
[2] *Assistant professor, Department of Information Technology, L.D. College of Engineering, Gujarat, India*

ABSTRACT

As e-commerce becomes more and more mature, number of reviews about products grows as well. For fine-grain opinion mining, opinion words and opinion targets extraction is important task, the main part of which is detecting different opinion relation among words. For this task, this paper proposes a method based on word alignment model, which also consider topical relation along with alignments. After finding alignments, the confidence of the candidate is estimated. Then candidate with higher confidence is extracted. Finally topical relations are considered in addition to extracted candidates. All previous methods are either dictionary based or based on nearest-neighbor rule, or based on NLP rules. Dictionary based methods are domain specific, nearest-neighbor rules based methods are not precise when long span relations and NLP rule based method not precise when used on informal inline texts. Our experiment results on two different data, for different domains shows that consideration of topical relation improve the method based on word alignment model.

Keywords: *E-commerce, opinion mining, fine-grain opinion mining, opinion target extraction, opinion word extraction, word alignment model, NLP rules, topical relation*

1. INRODUCTION

Now-a-days e-commerce technology become more popular, the convenience of online shopping has attracted more and more people. In order to get product feedback timely and to update future customer with other's shopping experiences of the same product, it is common for merchants to allow their customers to leave product reviews. As the number of customers increases, the number of reviews about the product grows as well. So mining of reviews become important process. Using opinion mining customer, who wants to buy the product, can have idea about the product's quality and manufacturer can improve the product on time [1].

Opinion mining task can be done using main two approaches: Sentiment classification and Feature based opinion mining [2]. In sentiment classification, the sentiment or opinion of the user is classified as positive or negative and then the summarization is done where as in feature based opinion mining, the feature is selected from review and the sentiment is extracted after feature selection.

There are three different levels of opinion mining; Document level, Sentence level and Entity and Aspect Level. The different levels of analysis are explained in details as follows [3]:

1. Document Level: The task at document-level analysis is to classify whether a whole document express a positive or negative opinion. This level of opinion mining assumes that each document gives opinion on a single entity (i.e. single product or single topic).
2. Sentence Level: In this type of analysis the task is to identify whether each sentence gives a positive, negative or neutral sentiment. Neutral usually means no opinion in the sentence. This level of analysis is related to subjective classification, which distinguishes objective sentences that expresses factual information from subjective sentences that expresses subjective views and opinions.
3. Entity and Aspect Level: Aspect level analysis performs fine-grain analysis. Aspect level is earlier called feature level. Instead of looking at language constructs (documents, sentences, paragraphs, etc.), aspect level directly look for opinion. It is based on idea that an opinion consists of a sentiment (positive or negative) and a target (of opinion).

For customer it is easier when opinion mining is fine-grain about a product feature. For example: "this phone has a colorful and big screen." Customer expect positive opinion of phone's screen. So, opinion words and opinion targets are must be detected. Which is considered in this paper.

An opinion target is defined as the object about which user expresses their opinions, typically noun or noun phrases. In above example "screen" is opinion target. Opinion words are defined as the words which are used to express users' opinion. In example given above, "colorful" and "big" are opinion words. Word alignment model is use to find alignment relation between words. Opinion relation among words is shown in figure 1 [7].

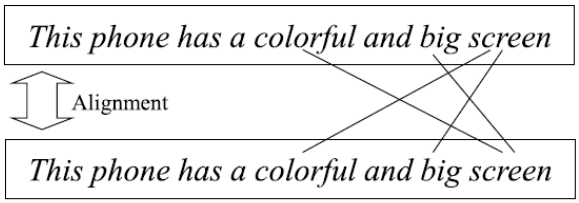

Fig -1: opinion relation between the words using word alignment model

2. RELATED WORK

Lots of work is done for opinion or review mining. There are many methods which have their own advantages and disadvantages. Opinion words and opinion targets are new work in the mining area. Some methods simply do sentiment analysis, some other do feature based analysis, and some use supervised approach.

Gaurav Dubey [5] proposed a simple method for sentiment opinion mining. The method uses the publicly available dataset of online product reviews by automatic extracting marketing intelligence from the vast repository of user generated opinions through main three steps: Part-Of-Speech (POS) tagging, Sentiment Analysis through rule mining, and summarizing and displaying the output. Part-of-Speech Tagging (POST) or lexical set is used to find out the grammatical words in any document or user speech: like noun, verb, adjective etc. [4]. This method is simplest method and gives overall review i.e. positive or negative about the product. This method does not give the review mining around the product feature. This is rule based method, so it cannot give good results when the reviews are in informal language.

LiZhen Liu [6] proposed fine-grain method to extract the product feature. In the process first of all LDA (Latent Dirichlet Allocation) model is used to extract the features that people are most interested in. After feature generation the strength of the opinion words are calculated. Using the strength, the emotional intensity is classified into six levels: Positive (strength, neutral, weak) and negative (strength, neutral, weak). Finally, all features are ranked according to the frequency of their appearances in the reviews. This method gives summarization of reviews around the product feature. So, the mining is more useful than Gaurav dubey's method. This approach gives summarization result in form of listing all comments.

Kang Liu [7] proposed a novel approach to extract opinion targets based on word based translation model (WTM). In this method opinion target extraction is composed of two main tasks: mining association between opinion target and opinion words using word based translation model and opinion words, candidate confidence estimation using Google n-gram corpus. At the end, the candidate with higher confidence than the threshold will be extracted as opinion target. In this system words with less occurrence, are ignored in word alignment model.

A. Jeyapriya [8] proposed supervised learning algorithm based method to extract aspects and mining opinions in product reviews. This method uses customer reviews to extract aspect and mine whether given review is positive or negative opinion. At first, the method removes Stop words. Stop words are words which are most frequently used in any language and not useful in text mining. Then, stemming is performed to form root word of a word. Here, Porter stemmer algorithm is used to stem words. After the stemming, POS tagging is performed. Stanford tagger is used for POS tagging. Finally, aspect extraction is performed. Nouns are extracted as aspect and then its frequency is compared to minimum support count. The word which has higher count than minimum support count is extracted. The Naïve Bayesian algorithm using supervised term counting based approach is used for sentence and aspect orientation. Finally, the system identifies the number of positive and negative opinions of each extracted aspect in customer reviews. The main disadvantage of this method is words whose occurrence is less will be not considered.

Kang Liu [9] proposed method for co-extracting opinion targets and opinion words from online reviews based on the word alignment model. In this method, the process of extracting opinion targets/words is viewed as co-ranking process. The opinion relation is found using word based alignment process. To get the optimal alignments an EM based algorithm is used. The standard alignment model is usually trained using unsupervised approach. So,

for improving alignment process, partial supervision is performed on the statistic model to incorporate partial alignment links into alignment process. After that opinion association among words is calculated. Then, candidate confidence estimation is done using random walking with restart algorithm and penalizing high degree vertices. At the end, candidate with higher confidence than minimum threshold is extracted as the opinion target or opinion word. This approach is most effective. But, the words with less occurrence are ignored by word alignment model. So, it further can be improved, which is the main focus this paper.

3. THE PROPOSED SYSTEM

In proposed system, the most effective technique i.e. based on word alignment model will be modified. In modification, we are considering topical relation in addition to word alignment relation. The method based on word alignment model uses IBM3 model for alignment process, which does not consider words that are occurred less time in input. So, some opinion targets and words are missed. Proposed system focus on those words. Here, topical relation, i.e. relation based on position is considered for covering missing opinion words and opinion target.

3.1 Overview of the proposed system

Main modules of the proposed system are shown in figure 2. Alignment links represents relation between potential opinion word and potential opinion target. Then, topical relations are found and both relation are merged. After that confidence of the candidate is estimated and candidate with higher confidence are extracted.

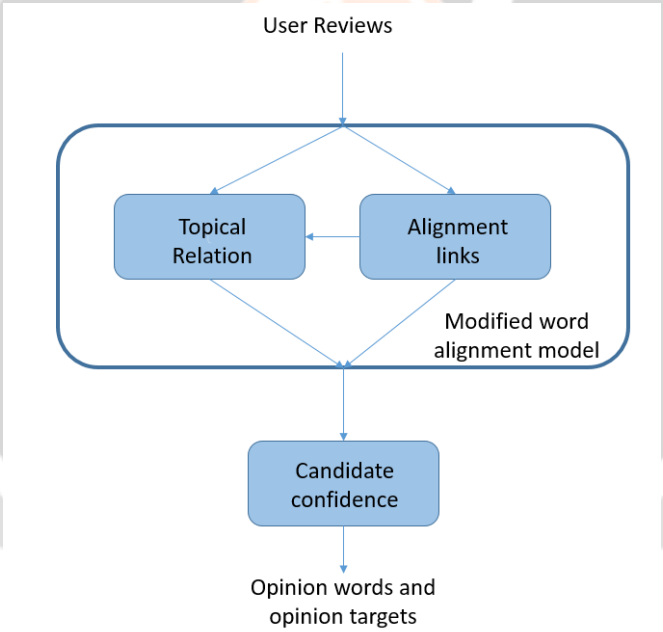

Fig -2: Main modules of the proposed system

3.2 Finding relations between opinion words and opinion targets

Opinion relation alignment links are found using IBM model 3 because it has been proven to perform better than any other model [9]. It uses bilingual word alignment algorithms which are here applied to monolingual input. We have

$$P_{ibm}(A|S) \propto \prod_{i=1}^{n} n(\emptyset_i, w_i) \prod_{j=1}^{n} t(w_j, w_{a_j})\, d(j|a_j, n) \quad 1)$$

Where $t(w_j|w_{a_j})$ is for modeling the co-occurrence information, $d(j||a_j, n)$ models word position information, $n(\emptyset_i|w_j)$ denotes the ability of word for "one-to-many" relation. Then hill climbing algorithm is applied to found alignments links. High-precision syntactic patters are fed to IBM model to make it partially

supervised. Fully supervised concept is impractical because it is impossible to manually label all alignments. After that to obtain optimal alignments EM based algorithm is used.

For capturing topical relation following algorithm is used. In the figure 3 n_{ft} is the number of the topical relation, $f_{k1,l1}$ is topical relation between k1 and l1, Â is optimal alignments found by hill climbing algorithm, ε_w is the threshold value for consideration of topical relation.

1. *while* $n_{ft} > 0$ *do*
2. *if* $f_{k1,l1} \in$ Â *then*
3. *do nothing*
4. *endif*
5. *else*
6. *if* $f_{k1}, l1 = \varepsilon_w$ *then*
7. Â U $f_{k1}, l1$
8. *endif*
9. *endelse*
10. $n_{ft} = n_{ft} - 1$
11. *end*

Fig -3: Algorithm for considering topical relation

3.3 Confidence estimation

Form the alignment result, alignment probability between a potential opinion target w_t and potential opinion word w_o is calculated using following equation.

$$P(w_t|w_o) = \frac{Count(w_t, w_o)}{Count(w_o)} \quad 2)$$

Similarly, we can find $P(w_o|w_t)$ by changing alignment direction in the alignment process. And then opinion association $OA(w_t, w_o)$ between w_t and $w_o$is calculated as follows.

$$OA(w_t,w_o) = \left(\frac{\alpha}{P(w_t|w_o)} + \frac{1-\alpha}{P(w_o|w_t)}\right)^{-1} \quad 3)$$

At the end confidence of the candidate is calculated using random walk with restart algorithm. Thus, we have

$$C_t^{k+1} = (1 - \mu)\, M_{to} \times c_o^k + \mu \times I_t$$

$$C_o^{k+1} = (1 - \mu)\, M_{to}^T \times c_k^k + \mu \times I_o \quad 4)$$

Where, C_t^{k+1} and C_o^{k+1} are confidence of opinion target and opinion word candidate, respectively $k+1$ iteration, C_t^k and C_o^k are confidence of opinion target and opinion word candidate, respectively in k iteratarion, M_{to} records opinion association among candidates, $m_{ij} \in M_{to}$ means opinion association between ith opinion target candidates and jth opinion word candidate, I_o and I_t are prior knowledge of candidate. Candidate with higher confidence are extracted as opinion target or opinion word.

3.4 Experiments and results

We have conducted experiments on the customer reviews of two different product Nokia 6610 and Canon G3. The reviews are collected from amazon.com data set. We can use review on any product from any website data, because the system is domain independent. Performance of the system is evaluated using three parameters: precision, recall and F-measure. Precision is measure of retrieved instances that are relevant. Recall is the fraction of relevant instances that are retrieved. F-measure is a measure of test's accuracy. To calculate these measures, true values in reviews are identified manually. The definitions of precision, recall and F-measure are as follows.

$$Precision = \frac{extracted\ values\ \cap\ true\ values}{extracted\ values}$$

$$Recall = \frac{extracted\ values\ \cap\ true\ values}{true\ values}$$

$$F - measure = \frac{2 * recall\ * precision}{recall\ + precision}$$

Considering topical relation with opinion relations gives better results than only opinion relation consideration as follows.

Table -1: Experimental results of opinion word extraction

Product name	With topical relation			Without topical relation		
	Precision	Recall	F-measure	Precision	Recall	F-measure
Nokia 6610	0.78	0.74	0.71	0.73	0.56	0.63
Canon G3	0.83	0.78	0.80	0.77	0.56	0.64
Average	0.81	0.76	0.78	0.75	0.56	0.64

Table -2: Experimental results of opinion target extraction

Product name	With topical relation			Without topical relation		
	Precision	Recall	F-measure	Precision	Recall	F-measure
Nokia 6610	0.80	0.82	0.81	0.80	0.73	0.76
Canon G3	0.82	0.84	0.83	0.76	0.81	0.78
Average	0.81	0.83	0.82	0.78	0.77	0.77

The comparison charts between two methods are shown in figure. Here, for comparison we use average values of parameters. The graph results clearly shows that consideration of topical relation improves the system.

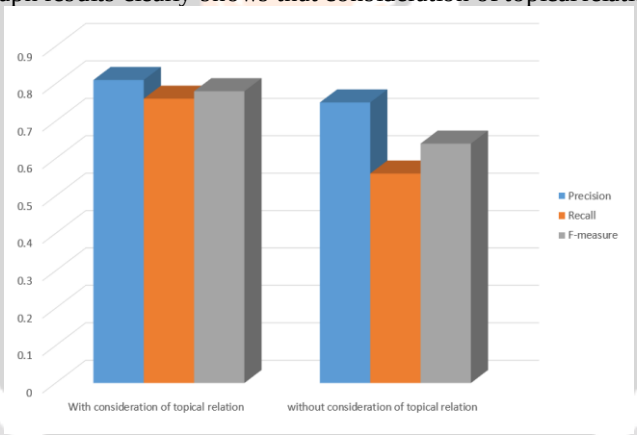

Chart -1: Parameter comparison for opinion word extraction

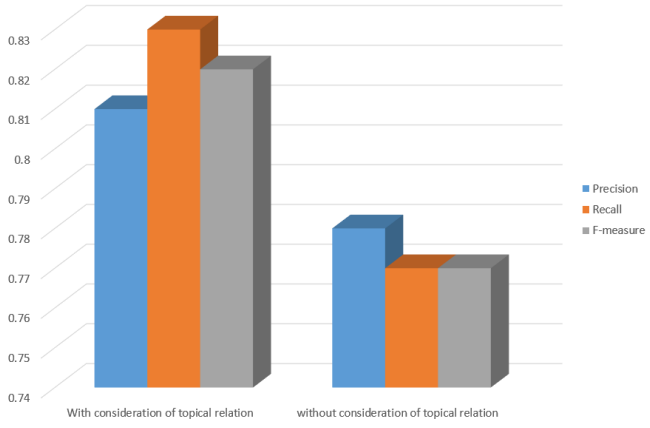

Chart -2: Parameter comparison for opinion target extraction

4. CONCLUSIONS

We have studied variety of methods, their advantages and disadvantages. The word alignment based method is most effective method for feature based opinion mining. So here we are trying to improve the most effective method. This paper proposes a method that co-extracts opinion words and opinion targets by using modified word alignment model with consideration of topical relation. Our main contribution is focus on the relations with are not covered using word alignment model. Form the result we can see that our system give better result than the word alignment model based method.

5. REFERENCES

[1]. Bing Liu, "Sentiment Analysis and Opinion Mining", Synthesis Lecture on Human Language Technologies, Morgan and Claypool Publishers, In May, 2012.
[2]. G. Angulakhmi, Dr. R. ManikaChezian, "An analysis on opinion mining; Techniques and tools", International journal of advanced research in computer and communication Engineering, In July, 2014
[3]. Vibhuti Patel, Mital Panchal, "A survey on Opinion Mining Methods from Online Reviews", International Journal of Scientific Research in Science, Engineering and technology, In December, 2015.
[4]. Bo Pang, Lillian Lee, "Opinion mining and sentiment analysis", Foundations and Trends in Information Retrieval, in 2008.
[5]. Gaurav Dubey, Ajay Rana, Naveen Kumar Shukla, "User Reviews Data Analysis using Opinion Mining on Web", In 2015, IEEE.
[6].LiZhen Liu, WenTao Wang, HangShi Wang, "Summarizing Customer Reviews Based On Product Feature", In 2012, IEEE.
[7]. Kang Liu, Liheng Xu, Jun Zhao, "Opinion Target Extraction Using Word-Based Translation Model", In 2012, Joint Conference Empirical Methods Natural Language Processing Computer Natural Language Learning.
[8]. A.Jeyapriya, C.S.KanimozhiSelvi, "Extracting Aspects and Mining Opinions in Product Reviews using Supervised Learning Algorithm", In 2015, IEEE.
[9].Kang Liu, Liheng Xu, Jun Zhaao, "Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on The Word Alignment Model", In 2015, IEEE Transaction.