RESEARCH ARTICLE OPEN ACCESS

Aspect-Based Opinion Mining Using Dependency Relations

Amani K Samha

Science and Engineering Faculty
Queensland University of technology
Brisbane 4000
Queensland Australia

ABSTRACT

Over the course of recent years, Opinion Mining from unstructured natural language text has received significant attention from the research community. In the context of Opinion Mining from customer reviews, machine-learning approaches have been recommended; however, it is still a very challenging task. In this paper, we have addressed the problem of Opinion Mining, and we propose a Natural Language Processing approach that undertakes Dependency Parsing, Pre-processing, Lemmatization, and part of speech tagging of natural texts in order to obtain the syntactic structure of sentences by means of a dependency relation rule. Specifically, we employ Stanford dependency relations and Natural Language Processing as linguistic features and present an Aspect-Based opinion mining extraction algorithm from customer reviews. Throughout this paper, we also highlight the importance of subjective clause lexicon. We evaluate our extraction approach using customer product reviews collected from Amazon for nine different products collected by Hu and Liu [1]. Based on empirical analysis, we found that the proposed dependency patterns provided a moderate increase in accurate results than the baseline models. This study also found that the average per cent change for aspect and opinion extraction was significantly improved compared to the baseline models. We show the results of our study and discuss how they relate to comparative experimental results. We end with a discussion that highlights the strong and weak points of this method, as well as direction for future work. Examples are provided to demonstrate the effectiveness of using Dependency Relations for optimizing the problem of Opinion Mining.

Keywords:- Opinion Mining, Sentiment Analysis, Dependency relations, Natural Language Processing.

I. INTRODUCTION

In the Web 2.0 platforms, enormous amounts of information are shared wherein people exchange their opinions and benefit from others' experiences. This includes social media, forums, blogs and product reviews. In fact, customer reviews have become a thrilling reference that is used in most industries, such as business, education and e-commerce. Most customer reviews contain opinionated information about a person's personal experience with certain services and products [2]. As a person analyses existing reviews for a certain product or service, his or her decision-making process is enhanced. In the business world, for example, reviews may help improve the way that services or products are offered by the seller to potential customers based on earlier customers' experiences and feedback. It may also influence the likelihood of someone who is simply browsing the site actually becoming a paying customer. It is clear that the decision-making process is highly enhanced by reviews on the business side and consumer side alike.

The ability to post opinionated reviews is a service that is provided by many e-commerce websites, such as eBay, Amazon and Yahoo Shopping, whereby customers can post their opinions as free text. Although the process seems straightforward, it involves a huge amount of work due the complexity of natural language and the number of reviews. For example, to go through all reviews and form an opinion based on them could be highly time consuming and difficult. Consequently, creating a system that gathers all of the information, analyses it, and extracts useful knowledge from it is very challenging. A successful system needs to offer the highest benefit at a minimal level of effort to all parties involved.

Opinion Mining, in general, is classified into three levels: the document level, which aims to provide an overall opinion; the sentence level, which produces opinions based on the sentence; and the feature level, which examines each feature in the review. Aspect-Based Opinion Mining (ABOM) is the core focus of this study, many other researchers such as [3-5]. Generally, ABOM involves several tasks. Firstly, it aims to efficiently identify and extract product entities from relevant reviews. This includes the actual product, its components, functionality, attributes and the aspects of the product [6]. Secondly, it finds the corresponding opinions for each entity extracted. Opinions are also known as 'sentiments', which are subjective and are presented as adjectives in the sentence to express how the customers feel about the product or the service. Finally, a summary of that information is presented, which is known as Opinion summary, and mostly contains the sentiment as well.

The requirements for comprehensive information are addressed by aspect-based opinion mining. Many approaches are recommended for extracting aspects from reviews. Several of such works utilized complete text reviews that contain irrelevant information, whereas others took benefits of short comments. Numerous algorithms are also provided to identify the aspects' rating. Estimated ratings and extracted aspects offer more comprehensive information to users for making decisions and to suppliers for monitoring their consumers [7].

Given a group of reviews regarding item P, the job is to identify the k key aspects of P and for predicting every aspect's rating. Two main broad tasks are involved in aspect-based opinion mining, starts with the aspect identification, then finding corresponding opinion and its orientation. This task aims at extracting aspects of the item reviewed and to group aspects' synonyms, for various people can use various phrases or words for referring to the aspect. For instance, display, LCD, screen, Rating prediction: The aim of this task is to determine whether opinion on the aspect is negative/positive or approximating the rating of the opinion in the range of 1 to 5[7]. "Google Shopping" 1, previously "Google Product Search", an internet marketplace was launched by Google Inc. Users may type product queries for returning lists of vendors marketing a specific product, and also pricing information, product general rating and reviews of product. In Google Shopping, product reviews are from sites of third party. For instance, digital camera's reviews are collected from ConsumerSearch.com. NewEgg.com, Epinions.com, BestBuy.com, etc. Furthermore, to list the review texts, Google Shopping applied the technique of aspect-based opinion mining for extracting aspects of product from reviews. Also it offers the percentages of negative and positive sentences for every aspect extracted for helping users in decision-making. A number of researchers have attempted to solve the Opinion Mining problem using different approaches via supervised, unsupervised and semi-supervised learning. These include rule-based methods [8-12], statistical methods [8, 13-15] and lexicon approaches [16, 17] [18-20]. In this paper, we study the problem of ABOM from a linguistic perspective, and propose an approach using Natural Language Processes (NLP) techniques along with subjective clauses lexicon of product reviews. Recent research has shown that the NLP techniques based on dependency relations actually enhance the accuracy and performance of unstructured prediction problems. The main contributions are the use of dependencies to find product features, such as opinion pairs employing subjectivity knowledge. We also measure the impact of using lemmatization processes from the beginning of pre-processing rather than at the end of the process. This paper is organized as follows: Section 1 introduces the background and significance of the study; Section 2 discuss the related work and motivation; Section 3 describes the Aspect-Opinion Mining Extraction Method based on dependency parsing; Section 4 evaluates the experimental results and analyse errors. Finally, the research is reviewed in a summary discussion and direction for future research is provided.

II. RELATED WORK

Various extraction methods have been proposed for Opinion Mining and Sentiment Analysis in unstructured text such as customer reviews [1, 9, 21-29]. Different levels of Opinion Mining can be a good source for providing an overall polarity of a whole document [11, 30-33] or sentence [9, 34, 35]. However, it fails to detect the sentiment relative to product aspects in a document or a sentence. Aspect-based Opinion Mining of product reviews, on the other hand, works by identifying opinion targets and mapping the opinion-bearing words by using domain based lexicon, similar to what Kanayama and Nasukawa [36], Kaji and Kitsuregawa [37, 38] have done. Nevertheless, most previous extraction methods mostly rely on part of speech (POS) tags and some syntactic information.

In this paper, we focus our study on aspect extraction at a sentence level using different NLP techniques. In the past, dependency patterns have been hypothetically employed in a variety of fields using different approaches to identify product aspects and their corresponding opinions from reviews in several languages. We highlighted the most recent approaches that share likenesses with our own approach. Different feature selections were used along with machine-learning, including unigrams and bigrams by, for instance, Pang et al. [31]. Meanwhile, Matsumoto et al. [39] use syntactic relations between words in sentences for document sentiment classification. Agrawal et al. [40] used dependency relation based on words to extract features from text based ConceptNet ontology, and then used the mRMR feature selection technique to element redundant information. Somprsertsri and Lalitrojwong [41] still propose a system that mines opinion and product aspects considering the syntactic and semantic information and is based on dependency relations and ontology knowledge. Kumar and Raghuveer [42] used the opinion expressions to find product aspects and identify opinionated sentences by proposing sematic rules. Popescu and Etziono [26] introduced the OPINE system, which uses syntactic patterns to mine the orientation of opinions based on unsupervised information extraction. There is similar work that has been done, but in different languages—such as Spanish—by Vilares et al. [43], where NLP techniques were combined with the syntactic structure of sentences to mine aspects and opinions and then find their orientation.

III. DEPENDENCY RELATION FOR ASPECT-OPINION

Our method consists of two major steps: the first is the pre-processing and the second is the main processing. Each step contains a sequence of sub steps where each used different language tools. At the beginning, we had to define what product aspects are: the aspects are anything related to the product, including the product itself and/or part components and functions of the product.

According to Banitaan et al. (2010) and Glance et al. (2004) the aspects can be classified under the entity definition and categories, as illustrated in Table I.

TABLE I ENTITY CATEGORIES

Entity	Description
Components	Physical objects of a camera, including the camera itself, the LCD, viewfinder and battery
Functions	Capabilities provided by a camera, including movie playback, zoom and autofocus
Features	Properties of components or functions, such as colour, speed, size, weight, and clarity
Opinions	Ideas and thoughts expressed by reviewers on the product, its features, components or functions
Other	Other possible entities defined by the domain

However, the broader consensus among researchers categorizes them into four entity groups that represent different types of words in the review text. These four categories are components, functions, features and opinions. For instance, Table I includes an example of entity categories related to the word 'camera' (Glance et al., 2004). In many cases, certain entities may not fit in to any of the four categories. Therefore, a fifth category is formed, called 'other', which is left open for certain suggested categories that do not belong to any of these four entity categories. We used the word "aspect" to present the actual product, components, function and features of the product. The core methods aspect and opinion extraction using dependency relation are described in the following subsequent sections.

In this section, we describe our proposed approach to mine product aspects from online customer reviews. The proposed approach is divided into two main correlated tasks. The first task is to prepare the dataset by employing NLP techniques. The second task is to find Opinion words and map them to the product aspects. Fig. 1 shows the architecture of the whole system while the subsequent section describes all the steps and provides explanatory examples.

A. Data Pre-processing

This is a review written by a customer for a camera. Here, it will be used as an example. *"[t] do not buy this piece of junk . ##i purchased this unit 3 months back and i think the unit knew when my warranty expires. Picture [-2], player[-3][p]##it is more than 90 days and it does not show the picture no matter what i do .##i can only hear the sound"*.

The pre-processing was accomplished using NLP techniques as follows: First, we clean up the dataset using regular expressions, where we remove symbols such as *{, [, :), :(…,* since the reviews are natural text and are full of unnecessary characters and abnormal symbols. Once the dataset has been cleaned out, this is how it appears:

- (do not buy this piece of junk. i purchased this unit 3 months back and i think the unit knew when my warranty expires. It is more than 90 days and it does not show the picture no matter what i do . i can only hear the sound).

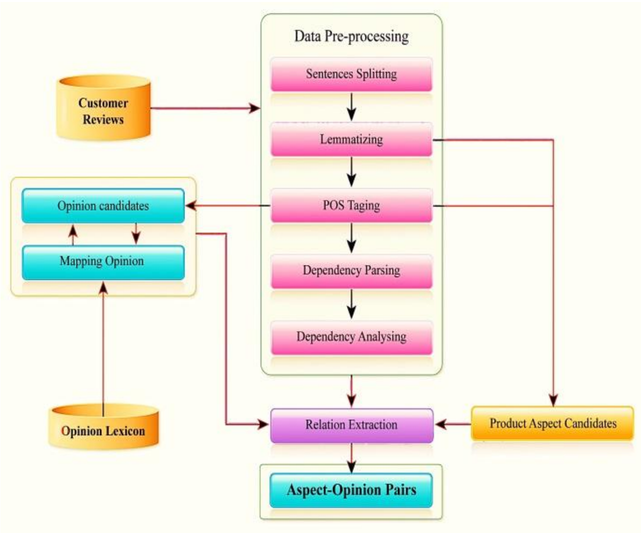

Fig. 1 System Architecture

After removing all symbols, we use Stanford Lemmatization. Determining the lemma tags for a given word allows us to change the form of a word so that all words can be treated as a single item for extraction reasons. We use lemmatization to prepare the text files, as it helps to find all possible aspects and group similarities based on the forms of a word. It works by removing the endings of words and returning the word to the base or dictionary form of the word, which allows us to group different, forms of words together as a single item. After this step, the dataset is ready for the next step (POS tagging). Previous research commonly performed lemma tagging at the end, but we did the lemma at the beginning of the process, aiming to group smiler words to find frequent aspects and opinions treat them as single item.

- "do not buy this piece of junk . i purchase this unit 3 month back and i think the unit knew when my warranty expire . it is more than 90 day and it does not show the picture no matter what i do . i can only hear the sound"

We then use Stanford Core NLP library version 3.4 annotators to split sentences, which allows us to work at a sentence level. By splitting sentences, we can draw the boundaries, which in turn let us continue working under the assumption that the aspects and its corresponding opinion can be found within the sentence boundaries. This is how the dataset appears after it has drawn the sentence boundaries:

- do not buy this piece of junk.
- i purchased this unit 3 months back and i think the unit knew when my warranty expires.
- It is more than 90 days and it does not show the picture no matter what i do.
- i can only hear the sound .

The next step is to run the POS tagging, in which we aim to find which part of speech each word is (such as verb, noun, adjective, etc.), and will help us fulfil the assumptions of this paper. Fig. **2** shows an example.

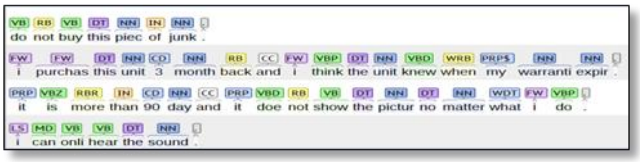

Fig. 2 POS Tagging Example

Finally, we run the Stanford Dependency Relations to find the syntactic parsers that will allow us to map the dependencies between all words within the sentence in the form of relation (governor, dependent). "The dependencies are all binary relations: a grammatical relation holds between a governor (also known as a regent or a head) and a dependent". All the grammatical representations, abbreviations are illustrated in[44] . For simplicity's sake, we will use one example, as seen in Fig. **3**. We will present one example and the same applied to all sentences. The Parse tree for the first example "do not buy this piece of junk." is illustrated in Fig. **4**.

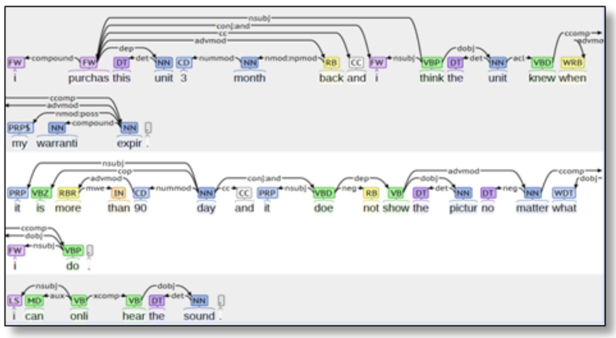

Fig. 3 Dependency Parser Example

```
aux(buy-3, do-1)
neg(buy-3, not-2)
root(ROOT-0, buy-3)
det(piece-5, this-4)
dobj(buy-3, piece-5)
case(junk-7, of-6)
nmod:of(piece-5, junk-7)
```

Fig. 4 Dependencies Relation illustration

B. Product Aspect and Opinion Extraction

The aspect- and opinion-extraction are two steps that are interconnected. Before we applied the dependency relations rules, we studied some rules based on observations and some rules from previous work by numerous researchers [40] [45] [42] [46]. We organized the next section as follows: first, we list the most useful dependency relations from previous work and our new dependency relations Table II (presented in bold); next, we discuss some related assumptions for aspect extraction and evaluate then, then, we validate the closest assumption (Table III and Fig. 5) and apply the best combination of dependency relations—illustrated in Table II—to extract aspects. Finally, we integrated the extracted aspects with the opinion lexicon to find the corresponding opinion for each aspect. All dependency explanations, abbreviations and acronyms can be found in [44].

TABLE II DEPENDENCY RELATIONS PATTERNS

Dependency #	Regular
1	**nsubj(OpinionADJ, TargetNOUN)**
2	nsubj(Opinion,Target2) nn(Target2, Target1)
3	nsubj(Opinion, H) xcomp(Opinion, W1) dobj(W1,Target2) nn(Target2,Target1)
4	nsubj(Opinion,H) dobj(O,Target)
5	nsubj(W1,Opinion) acomp(W1,Target)
6	nsubj (W1, H) acomp (W1, Opinion) rcmod(Target2, W1) nn(Target2,Target1)
7	amod(Target, W1) amod(W1, Opinion)
8	amod(Target, W1) conjand(W1, Opinion)
9	amod(W1,Opinion) conjand(W1,Target)
10	nsubj(Opinion,H) prepwith(O,Target2) nn(Target2,Target1)
11	nsubj(Target,Opinion2) nn(Opinion1,Opinion2)
12	amod(Target2, Opinion) conjand(Target2,Target4) nn(Target4,Target3) conjand(Target2,Target5) nn(Target2,Target1)
13	**amod(TargetNOUN, OpinionADJ)**
14	**nmod(OpinionADJ, TargetNOUN)**
15	**nmod(W, TargetNOUN) nsubj(W, OpinionADJ)**
16	**xcomp(W,Opinion) nsubj(W,TargetNOUN)**

After the aforementioned steps, we then considered some of the assumptions regarding aspect extraction, and we evaluated them. Some research [1, 9] [40] [45] [42] [46] assumed that nouns could be listed as aspect candidatures. We applied this assumption to our dataset. Most of the aspects are highly relevant by assuming that all words can be aspects candidatures A_1 by a percentage of 94%. However, this assumption will not be considered since the dataset contained stopping words. Other words are candidatures for opinions and some other words are neither aspects nor opinions.

Fig. 5 shows that an A5 is a balanced assumption, therefore, we assume that the most frequent nouns and adjectives are aspect candidatures; however, there another assumption is that most opinions are adjectives. From this point on, we needed to apply another assumption to validate the initial assumption. We tested all proposed rules based on each assumption (A1 to A6), without the pre-processing to verify our approach. Examining all the above rules along with the aspect initial assumptions, led us to the perfect combination of syntactic rules that achieved high accuracy compared to the baseline model, which will be discussed in the results section.

Aspect ID	Aspect Technique	Precision	Recall	F-measure
A_1	All words as aspects (unigrams)	0.142	0.949	0.247
A_2	All nouns as aspects	0.046	0.758	0.088
A_3	Most frequent nouns (50%)	0.340	0.563	0.424
A_4	All words as aspects (bigrams)	0.0	0.0	0.0
A_5	All nouns + adjectives as aspects	0.038	0.875	0.074
A_6	Most frequent nouns and adjectives (50%) as aspects	0.296	0.675	0.411

TABLE III ASPECT EXTRACTION ASSUMPTIONS

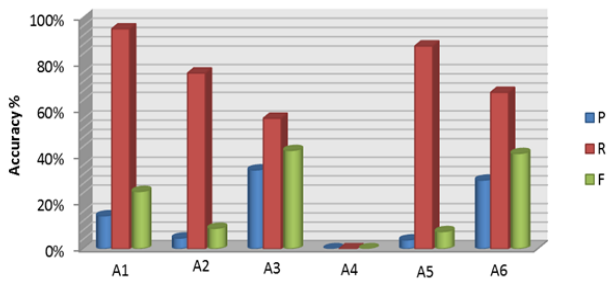

Fig. 5 Aspect assumption accuracy

In the subj dependency, if the POS tag of the governor is noun and the POS of the dependent is adjective, then we extract the opinion as the governor and the aspect as the dependent. In the mod dependency, we extract the opinion as the dependent, and the aspect as the governor, only if the conj_and dependency exist, correspondingly the next aspect is obtained from dependent and the same opinion is used. If an obj dependency exists, where the governor POS is not a verb then the opinion is the governor, we consider the next word as aspect in all cases the aspect. In the subj dependency exist, then the dependent is the opinion word, likewise in the comp dependency, whereas if the subj dependency exists, then the dependent is the aspect word. For all rules, we apply two other relations: nn or compound dependencies in order to find several aspects referring to the same opinion, also for orientation evaluation we apply the neg dependency relation. Then we applied Apriori Algorithm [47] with minimum support of 1% to find the most frequent product aspects list (FF). Finally, we merged all aspects, the FF list and the CFOP set, and then we mapped the relations with the Opinion lexicon, to generate the (FPF).

IV. EXPERIMENT RESULTS AND DISCUSSION

As mentioned earlier, our proposed approach has two main steps: pre-processing and aspect-opinion extraction. For the first step, we used Stanford CoreNLP, which includes POS tagger, lemmatization and syntactic parsing.

C. Dataset

For evaluation reasons we used two datasets; the first dataset is a subset of the second dataset. The two datasets that are involved in this research consist of annotated customer reviews of nine and five different products, respectively, collected from Amazom.com. Both datasets were collected and processed by Bing Liu [9] and [26] and contain approximately 4500 sentences, of which each dataset is about one product and consists of a minimum of 230 sentences written by customers as opinionated reviews. They were written as unstructured text files from a total number of 852 writers.

Based on the fact that opinions tend to be subjective, we decided to use subjectivity clauses that were represented in [23] as an opinion lexicon. Originally it was collected by [48] and was expanded using General Inquirer [49]. It contains positive and negative words with a total of over 8,000 subjective words and phrases. Then, the lexicon was categorized based on strength and weakness (StrongSubj or WeakSubj). Combining both dictionaries increased the accuracy of opinion extracting.

```
Algorithm 1: Aspect Based Opinion mining
  Input: D- Set of reviews. OL- Opinion Lexicon.
  Output: AO- Set of Aspects and Opinions, FPF- Set of Product Aspects
  Assumptions:
  AO = ∅;
  A = ∅;
  FF = ∅; Frequent Aspects;
  Method:
  Load NegList from negative-words.txt;
  Load PosList from positive-words.txt;
  foreach d ∈ D do
    Get set of sentences S from d ;
    foreach s ∈ S do
      Annotate each w ∈ s with Lemma, POS, Syntactic tag (NP);
      Extract all dependencies in ListDep for sentence s ;
      foreach G =< rel, gov, dep >∈ ListDep do
        if (G.rel = NSUBJ) then
          PairsExtract_NSUBJ(AO, G, ListDep, NegList, PosList)
        else
          if (G.rel = AMOD) then
            PairsExtract_AMOD(AO, G, ListDep, NegList, PosList)
          else
            if (G.rel = PREP_WITH) then
              PairsExtract_PREP(AO, G, ListDep, NegList, PosList)
            else
              if (G.rel = NMOD) then
                PairsExtract_NMOD(AO, G, ListDep, NegList, PosList)
              else
                if (G.rel = XCOMP) then
                  PairsExtract_XCOMP(AO, G, ListDep, NegList, PosList)
  FF=Apply Apriori Algorithm with minimum support 1% to obtain frequently itemsets ;
  A = A ∪ FF ;
  foreach ao =< aspect, opinion >∈ AO do
    if (ao.aspect ∉ FF) then
      FPF = FPF ∪ ao.aspect
  return AO, FPF
```

Fig. 6 Aspect-based Opinion mining algorithm

D. Evaluation criteria

To evaluate the efficiency of this research, different measures were used, namely: precision, recall, F-measure and percentage of change. In our experiment, we have a collection of documents, and every document has reviews related to a specific product. We used the aforementioned three measures to evaluate the relevance and irrelevance of the extracted features. Precision p is the fraction of the retrieved documents that is relevant to the topic, while recall r is the fraction of the relevant documents that has been retrieved. Those measures were discussed in further detail in [50], and they were calculated using the confusion matrix terms as shown in Table IV [50].

TABLE IV EVALUATION MATRIX

	Observation	
Expectation	TP (true positive)	FP (false positive)
	FN (false negative)	TN (true negative)

$$Precision(P) = \frac{TP}{TP + FP}$$

$$Recall(R) = \frac{TP}{TP + FN}$$

The TP is the number of positive documents, which means the relevant documents that are identified by the system. FP is the number of negative documents that are not relevant. FN is the number of relevant documents that the system failed to identify [50]. F-measure is another measure used to judge accuracy. It is calculated based on the precision and recall measures. The relationship between the value of the F-measure and the value of precision and recall is a direct relationship. Hence, if the value of precision and recall is high, the value of F-measure will also be high. The F-measure is calculated as follows [50]:

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

Finally, we used the percentage of change (PC) as an indicator of a change obtained from the new approach. PC is a ratio that is expressed as a fraction of 100.

$$\text{Percentage change PC} = \frac{N_{\mathrm{final}} - N_{\mathrm{initial}}}{N_{\mathrm{initial}}} * 100$$

E. Result Analysis

In this section, we analyse the results obtained from the employed dataset that we used to develop and evaluate our approach. We show a comparison of performance obtained for our system, and other approaches on aspect-based opinion mining from customer reviews. Given that our approach relies on rules, we therefore compared it to a state of the art system, which uses dependency relations. We used four accuracy measurements: precision, recall, F-measure and, finally, the percentage change. Precision and recall measures the retrieved and relevant aspects and opinions, and F-measure is the harmonic mean between Precision and recall. We used the percentage change as a way to evaluate the change in a variable, where it represents the relative changes between the baseline values and the new obtained values.

The percentage of change measures shows an increase in the aspect-based extraction, in which we scored an average increase of 23 % in precision 16 % in recall and 20% in f-measure compared to the baseline [42] as illustrated in

Table V and Fig. **7**. Likewise, the percentage is higher in the opinion extraction as well, in which we score 12% in precision 24 % in recall and 18% f-measure compared to the baseline [42] as illustrated Table VI and Fig. **8**.

V. ERROR ANALYSIS

In any natural language processing system, errors can happen due to the nature of the used datasets. For example: Reviews are written in an unstructured format; therefore, there are some spelling mistakes, which will directly result in not getting the correct syntactic dependency.

Table V Aspect extraction results

Aspect extraction														
Products	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	AVG	PC
P Baseline	55	61	61	62	64	65	68	70	70	73	75	75	76	67%	**23%**
P proposed	66	67	67	70	72	76	80	90	99	99	99	99	99	83%	
R Baseline	58	60	63	69	74	78	80	80	82	80	83	83	83	75%	**16%**
R proposed	69	70	72	82	87	92	92	93	93	94	94	95	99	87%	
F Baseline	56	60	62	65	69	71	74	75	76	76	79	79	79	71%	**20%**
F Proposed	67	68	69	76	79	83	86	91	96	96	96	97	99	85%	

TABLE VI OPINION EXTRACTION RESULTS

Opinion extraction															
Products	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	AVG	PC
P Baseline	44	44	53	54	58	60	66	67	80	83	83	84	84	66%	**12%**
P proposed	46	56	59	60	66	69	71	72	84	90	90	99	99	74%	
R Baseline	33	43	52	57	71	71	72	73	74	74	75	75	75	65%	**24%**
R proposed	50	61	65	72	77	77	83	85	81	93	93	95	99	79%	
F Baseline	38	43	52	55	64	65	69	70	77	78	79	79	79	65%	**18%**
F Proposed	48	58	62	65	71	73	77	78	82	91	91	97	99	76%	

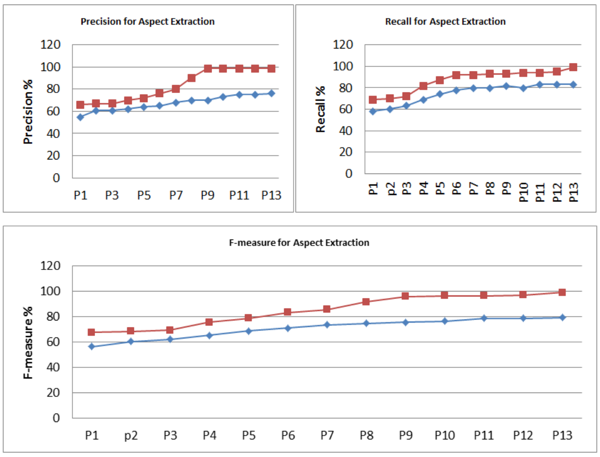

Fig. 7 Aspect Extraction Evaluation

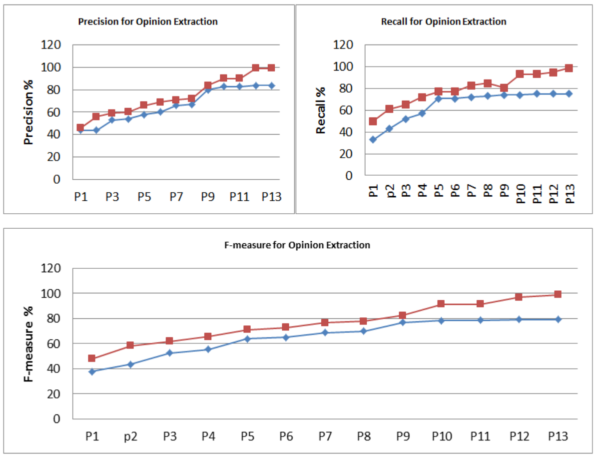

Fig. 8 Opinion Extraction Evaluation

With some comparative sentences, it is not easy to map the right relations between Opinions and product aspects; For example, "The picture quality of camera A is better than B", in which the opinion belongs to camera A not B.

As we have two datasets, the tables and graphs show the results from implementing our new approach on 13 different products.

Table Vshows results of product aspects extraction compared to the baseline model. Table VI shows the results of opinion extraction. The average precisions are 83% and 74%, respectively, and the average recalls are 87% and 79%, respectively. Fig. 7 and Fig. 8 show an increase in all performance measures; consequently, the consistent results prove the validity of our proposed approach compared to the baseline model.

REFERENCES

VI. CONCLUSIONS

The performance of the Opinion Mining and Sentiment Analysis process critically depends on the effectiveness of the aspect extraction process. Product reviews are a very valuable source for better purchasing and reselling decisions; however, posting enormous amount of reviews makes it hard to find useful information. A consideration the differences and preferences among consumers leads to the need to analyze the reviews in order to find all product aspects. The outcome, essentially, is to provide a better understanding to customers before buying. The study of aspect-based opinion mining has taken a preliminary step towards achieving this goal.

In this study, we proposed an approach to mine customer reviews and produce an aspect-based opinion-mining summary using dependency relations and subjective lexicon. Many product reviewers were analyses in order to glean an understanding of customer sentiment toward a product's attributes; opinion mining using different dependency rules was used in order to extract relevant information. Results related to our proposed approach were better than those that were obtained from a rules based approach [1, 9] and syntactic rules [42]. Therefore, we applied our approach to two different datasets. In both datasets, the accuracy was higher than in the baseline model. Consequently, we can say that our approach can be generalizable to different datasets. However, the improvement of subjective lexicon may reflect further improvement in the opinion extraction.

In summary, this paper proposed an aspect opinion mining approach for mining product aspects and corresponding opinions from customer reviews. Our approach incorporates subjective clauses lexicon and map relations using dependency relations of sentences. We explored a rich set of syntactic rules and relations that were observed from the product dataset and that demonstrated their effectiveness in the mapping of the relationships between the product aspects and the corresponding opinions. Our experiments showed that our model achieves better accuracy than existing dependency models for aspect-based opinion mining from customer reviews. Lastly, our approach for aspect-opinion relation extraction can be further improved by applying more rules. As a possible direction for future work, we might consider finding more useful dependencies along with expanding the Opinion Lexicon.

[1] M. Hu and B. Liu, "Mining opinion features in customer reviews," 2004, pp. 755-760.

[2] S. Moghaddam, M. Jamali, and M. Ester, "Review recommendation: personalized prediction of the quality of online reviews," in *Proceedings of the 20th ACM international conference on Information and knowledge management*, 2011, pp. 2249-2252.

[3] M. Himmat and N. Salim, "Survey on Product Review Sentiment Classification and Analysis Challenges," in *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*. vol. 285, T. Herawan, M. M. Deris, and J. Abawajy, Eds., ed: Springer Singapore, 2014, pp. 213-222.

[4] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, pp. 1-167, 2012.

[5] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," *Mining Text Data*, pp. 415-463, 2012.

[6] X. Ding, B. Liu, and L. Zhang, "Entity discovery and assignment for opinion mining applications," presented at the Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France, 2009.

[7] V. Bhatnagar, "Data Mining and Analysis in the Engineering Field " pp. 11-43, 2014.

[8] H. Guo, H. Zhu, Z. Guo, X. Zhang, and Z. Su, "Product feature categorization with multilevel latent semantic association," in *Proceedings of the 18th ACM conference on Information and knowledge management*, 2009, pp. 1087-1096.

[9] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168-177.

[10] M. Hu and B. Liu, "Mining opinion features in customer reviews," in *AAAI*, 2004, pp. 755-760.

[11] B. Liu, M. Hu, and J. Cheng, "Opinion observer: analyzing and comparing opinions on the web," in *Proceedings of the 14th international conference on World Wide Web*, 2005, pp. 342-351.

[12] S. Moghaddam and M. Ester, "Opinion digger: an unsupervised opinion miner from unstructured product reviews," in *Proceedings of the 19th ACM international conference on Information and knowledge management*, 2010, pp. 1825-1828.

[13] H. Wang, Y. Lu, and C. Zhai, "Latent aspect rating analysis on review text data: a rating regression approach," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2010, pp. 783-792.

[14] Y. Choi and C. Cardie, "Hierarchical sequential learning for extracting opinions and their attributes," in *Proceedings of the ACL 2010 conference short papers*, 2010, pp. 269-274.

[15] I. Titov and R. McDonald, "Modeling online reviews with multi-grain topic models," in *Proceedings of the 17th international conference on World Wide Web*, 2008, pp. 111-120.

[16] L. Zhao and C. Li, "Ontology based opinion mining for movie reviews," *Knowledge Science, Engineering and Management,* pp. 204-214, 2009.

[17] N. F. Noy, "Semantic integration: a survey of ontology-based approaches," *SIGMOD Rec.*, vol. 33, pp. 65-70, 2004.

[18] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexiconbased and learning-based methods for twitter sentiment analysis," *HP Laboratories, Technical Report HPL-2011,* vol. 89, 2011.

[19] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational linguistics,* vol. 37, pp. 267-307, 2011.

[20] F. Wogenstein, J. Drescher, D. Reinel, S. Rill, and J. Scheidt, "Evaluation of an algorithm for aspect-based opinion mining using a lexicon-based approach," in *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, 2013, p. 5.

[21] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and trends in information retrieval,* vol. 2, pp. 1-135, 2008.

[22] B. Liu, "Opinion mining and sentiment analysis," in *Web Data Mining*, ed: Springer, 2011, pp. 459-526.

[23] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in *Proceedings of the conference on human language technology and empirical methods in natural language processing*, 2005, pp. 347-354.

[24] M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger, "Pulse: Mining customer opinions from free text," in *Advances in Intelligent Data Analysis VI*, ed: Springer, 2005, pp. 121-132.

[25] G. Vinodhini and R. Chandrasekaran, "Sentiment analysis and opinion mining: a survey," *International Journal*, vol. 2, 2012.

[26] A. M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," 2005, pp. 339-346.

[27] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *Natural language processing and text mining*, ed: Springer, 2007, pp. 9-28.

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

[29] A. K. Samha, Y. Li, and J. Zhang, "Aspect-based opinion mining from product reviews using conditional random fields," 2015.

[30] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," 2002, pp. 417-424.

[31] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, 2002, pp. 79-86.

[32] L. Yu, J. Ma, S. Tsuchiya, and F. Ren, "Opinion mining: A study on semantic orientation analysis for online document," in *Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on*, 2008, pp. 4548-4552.

[33] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in *Proceedings of the 19th ACM international conference on Information and knowledge management*, 2010, pp. 939-948.

[34] J. Wiebe, T. Wilson, R. Bruce, M. Bell, and M. Martin, "Learning subjective language," *Computational linguistics*, vol. 30, pp. 277-308, 2004.

[35] S. Moghaddam and M. Ester, "AQA: aspect-based opinion question answering," in *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, 2011, pp. 89-96.

[36] H. Kanayama and T. Nasukawa, "Fully automatic lexicon expansion for domain-oriented sentiment analysis," in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, 2006, pp. 355-363.

[37] N. Kaji and M. Kitsuregawa, "Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents," in *EMNLP-CoNLL*, 2007, pp. 1075-1083.

[38] G. Qiu, B. Liu, J. Bu, and C. Chen, "Expanding Domain Sentiment Lexicon through Double Propagation," in *IJCAI*, 2009, pp. 1199-1204.

[39] S. Matsumoto, H. Takamura, and M. Okumura, "Sentiment classification using word sub-sequences and dependency sub-trees," in *Advances in Knowledge Discovery and Data Mining*, ed: Springer, 2005, pp. 301-311.

[40] B. Agarwal, S. Poria, N. Mittal, A. Gelbukh, and A. Hussain, "Concept-level sentiment analysis with dependency-based semantic parsing: A novel approach," *Cognitive Computation*, pp. 1-13, 2015.

[41] G. Somprasertsri and P. Lalitrojwong, "Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization," *J. UCS*, vol. 16, pp. 938-955, 2010.

[42] V. R. Kumar and K. Raghuveer, "Dependency driven semantic approach to product features extraction and summarization using customer reviews," in *Advances in Computing and Information Technology*, ed: Springer, 2013, pp. 225-238.

[43] D. Vilares, M. A. Alonso, and C. Gomez-Rodriguez, "A syntactic approach for opinion mining on Spanish reviews," *Natural Language Engineering*, vol. 21, pp. 139-163, 2015.

[44] M.-C. De Marneffe and C. D. Manning, "Stanford typed dependencies manual," Technical report, Stanford University2008.

[45] T. Chinsha and S. Joseph, "A syntactic approach for aspect based opinion mining," in *Semantic Computing (ICSC), 2015 IEEE International Conference on*, 2015, pp. 24-31.

[46] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," *Computational linguistics*, vol. 37, pp. 9-27, 2011.

[47] Y. Ye and C.-C. Chiang, "A parallel apriori algorithm for frequent itemsets mining," in *Software Engineering Research, Management and Applications, 2006. Fourth International Conference on*, 2006, pp. 87-94.

[48] E. Riloff and J. Wiebe, "Learning extraction patterns for subjective expressions," in *Proceedings of the 2003 conference on Empirical methods in natural language processing*, 2003, pp. 105-112.

[49] P. J. Stone and E. B. Hunt, "A computer approach to content analysis: studies using the General Inquirer system," presented at the Proceedings of the May 21-23, 1963, spring joint computer conference, Detroit, Michigan, 1963.

[50] B. Liu, *Web data mining: exploring hyperlinks, contents, and usage data*: Springer Verlag, 2011.