Identifying Opinion Mining Elements Based on Dependency Relations and Fuzzy Logic

Abdelaziz Bouras†, Haiqing Zhang*, Aicha Sekhari*, Yacine Ouzrout*
†ictQATAR Chair, Computer Science & Engineering Department, Qatar University,
PO Box. 2713, Doha, Qatar
*DISP laboratory, University Lumière Lyon 2, France, 160 Bd de l'Université 69676 Bron Cedex
abdelaziz.bouras@qu.edu.qa

Abstract: *Opinion mining mainly involves three elements: feature and feature-of relations, opinion expressions and the related opinion attributes (e.g. Polarity), and feature-opinion relations. Although many works are emerged to achieve the aim of gaining information, the previous researches typically handle each of the three elements in isolation that cannot give the sufficient information extraction results and hence increases the complexity and the running time of information extraction. In this paper, we propose an opinion mining extraction algorithm to jointly discover the main opinion mining elements. Specifically, the algorithm automatically builds kernels to combine closest words into new terms from word level to phrase level based on dependency relations, and we ensure the accuracy of opinion expressions and polarity based on fuzzy measurements, opinion degree intensifiers, and opinion patterns. The analyzed reviews show that the proposed algorithm can effectively identify the main elements simultaneously and outperform the baseline methods.*

Keywords: opinion mining; dependency relations; fuzzy sets and logic; opinion degree intensifiers; feature-by-feature analysis

1 Introduction

The widely used Web communication on mobile and web-based technologies has dramatically changed the way individuals and communities express their opinions. More and more reviews are posted online to describe customers' opinions on various types of products. These reviews are fundamental information to support both firms and customers to make good decisions. The features and attributes of a product extracted from online customer reviews can be used in recognizing the strengths and weaknesses of the heterogeneous products for firms. While customers do not always have the ability to wisely choose among a variety of products in the market, they commonly seek product information from online reviews before purchasing a new product.

Identifying the opinions in a large-scale document of customer reviews is an opinion mining issue, which is a sub-division of information extraction that is concerned with the features, with the opinion it expresses. Two fundamental problems of mining such information are opinion features extraction and opinion words locating.

Opinion features are characteristics of the products on which opinion has been described. Two issues are generated in product feature extraction. One is that synonyms are often occurring in extraction of features. The other one is some product features are combined by several nouns. Hence, feature-of relation is used to record the synonyms of features and rebuild the noun terms to more accurately represent product features.

Opinion expressions are the opinion words that the reviewers have adopted to describe their opinions on the related features. Opinion expressions are commonly composed by an opinion pattern involving adjectives, adverbs, and verbs instead of a single opinion word. Thus, opinion-of relation extraction is adopted to keep the opinion patterns. Opinion expressions also need to express the evaluation for correct targets. The feature-opinion relation extraction is necessary to be proposed to express the opinion expressions corresponding with the related opinion features.

This paper aims to solve the information extraction issues. And the remainder of this paper is organized as follows. Section 2 reviews the main related works. Section 3 represents the mechanism of assigning the polarity and intensity of opinion expressions. Section 4 introduces the opinion mining algorithm to jointly complete opinion mining tasks. Section 5 conducts the experiments in multiple aspects and analyzes the superiorities and the deficiencies of the proposed algorithm by comparing with the baseline works. Section 6 concludes the paper with future works.

2 Related work

In this paper, we focus on jointly detecting the three principle elements in the reviews: feature and feature-of relation, opinion and opinion pattern extraction, and feature-opinion extraction. In previous works, these elements have mostly been studied in isolation. Therefore, we treat these three elements as three separate tasks and study the related works.

The existing works on feature extraction can be divided into three groups: frequent term mining, supervised sequence labeling, and unsupervised and knowledge-learning based approach. The most representative work for "frequent term mining" approach is Hu and Liu (2004), which is restricted to detecting the features that are strongly associated with a single noun and considering only adjectives collocated with the near feature words as opinion expressions. Some additional works (Zhuang et al., 2006; Qiu et al., 2011) involve manually constructed rules and semantic analysis, but these still cannot fully reduce the disadvantages of this branch. The "supervised sequence labeling" (Jakob and Gurevych, 2010; Choi and Cardie, 2010) usually needs a large amount of training data that are mainly composed by hand-labeled training sentences. All of the methods mentioned above do not have the ability to group semantically related expression aspects together. The existing works belonging to "unsupervised and know-learning based approach (topic modeling) are based on two models: PLSA (Probabilistic Latent Semantic Analysis) (Hofmann, 1999) and LDA (Latent Dirichlet Allocation) (Blei et al., 2003). According to the work(Titov and McDonald, 2008), the existing models are not suitable to be used to detect features, because they can only work well for capturing global topics, but cannot intelligently understand human judgments.

The opinion expressions consist of a set of opinion words, which are used to present the polarities of sentiments and measure the strength of the expressed opinions. Previous research can be divided into two categories: CRF (Conditional Random Field)-based approaches and parsing-based approaches. Most of the CRF-based approaches mainly focus on one direction and single word expressions. However, all of the approaches belonging to this category are token-level and cannot efficiently extract phrase-level information. Although semi-CRFs (Okanohara et al., 2006) are proposed to allow sequence labeling in phrase-level, these methods are known to be difficult to implement (Yang and Cardie, 2012). Previous works(Kobayashi et al., 2007; Joshi and Penstein-Rosé, 2009) show that adopting syntactic parsing features to identify opinion expressions and the related attributes is more helpful than the CRF-based approach. Moreover, some combination approaches (Brody and Elhadad, 2010; Kobayashi et al., 2007) are proposed by considering the impacts between some internal elements.

In conclusion, all of the approaches have their own advantages and disadvantages. Although some models obviously outperform others in each element, to the best of our knowledge, there is no solution that is simultaneous proficient in all three elements in practice**.** In the opinion mining processes, the three elements usually lie in a labyrinth of relationships and one element will encounter another element in each sentence, which makes the opinion mining results not as straightforward to obtain. To be able to gain more benefits from actual practice for firms and customers, we aim to find a compromising solution that allows the three elements to be taken into account as an integrated unit instead of seeking the best approach for one element.

3 Fuzzy weights assigning for opinion expressions

Around 6800 positive and negative English opinion words were compiled by Hu and Liu (2004). We have extended these opinion words by adding some words that can express the degree of intensity in the customer's emotion. We have collected 62 adverbs that are called ***Opinion Degree Intensifiers,*** which can be used in both a positive and negative situation to express the opinion degree or to change the orientation of the opinion. Opinion Degree Intensifiers are grouped into two types: adverbs that only change the opinion degree; and adverbs that will change the orientation of the opinion. The opinion expressions have the characteristics of uncertainty as different customers will adopt different words to express the same opinion and the same word has different opinion intensity under different circumstances. Fuzzy logic is a sophisticated approach to tackle uncertain and inaccurate issues (Zhang et al., 2014). Therefore, five fuzzy degrees are defined for the first type of words based on the intensity of the adverbs. Three fuzzy degrees are given for the second type of words, because there are fewer words that have such function and the gaps among these words are narrow.

The 6800 opinion words are updated with assigned weights that lie in [-1, 1]. The sets of opinion words are categorized into five levels based on the orientation of the word. Some words are defined as the benchmark (core), which can be used as the standard when determining the other words' polarities.

To be able to know the fuzzy weights of every reviewer, two different cases are defined based on different combinations of opinion words in the proposed patterns.

Definition 3.1.1 (weights of case 1) The opinion is the combination of the opinion degree intensifiers and 6800 opinion words that include adjectives and verbs. The weights of the opinion in case 1 are defined in four types of situations based on the words' orientation, which is shown in the following equation:

$$weights\ of \begin{bmatrix} (\text{RB/RBR/RBS})\, combination \\ with\ (\text{JJ/RB/RBR/RBS}) \end{bmatrix} =$$

$$\begin{cases} 1.\ \text{degree}(\text{RB/ RBR/ RBS}) \oplus \text{degree(JJ/ RB/ RBR/ RBS)}, \\ if\ \text{degree}(\text{RB/ RBR/ RBS}) > 0\ \&\ \text{degree(JJ/RB/RBR/RBS)>0} \\ \quad \text{eg: very good, extremely high} \\ elseif\ \text{degree}(\text{RB/ RBR/ RBS}) < 0\ \&\ \text{degree(JJ/RB/RBR/RBS)>0} \\ \text{eg: not good, not high} \\ 2.\ -\text{degree}(\text{RB/ RBR/ RBS}) \oplus \text{degree(JJ/ RB/ RBR/ RBS)}, \\ if\ \text{degree}(\text{RB/ RBR/ RBS}) > 0\ \&\ \text{degree(JJ/RB/RBR/RBS)<0} \\ \quad \text{eg: very bad, extremely annoyed} \\ 3.\ \text{degree}(\text{RB/ RBR/ RBS}) \otimes \text{degree(JJ/ RB/ RBR/ RBS))}, \\ if\ \text{degree}(\text{RB/ RBR/ RBS}) < 0\ \&\ \text{degree(JJ/RB/RBR/RBS)<0} \\ \quad \text{eg : not bad, not annoyed} \end{cases} \quad (1)$$

Definition 3.1.2 (weights of case 2) Some opinion words appear together with case 1. For instance, "not a very good camera", "extremely high quality", etc. The opinion phrases of such types are calculated by Eq.2 and Eq.3.

$$weights\ of \left[(not\ /\ never\ /\ ...) \begin{bmatrix} combination\ with \\ (\text{RB/RBR/RBS}) \\ combination\ with \\ (\text{JJ/RB/RBR/RBS}) \end{bmatrix} \right]$$

$$= \begin{cases} 1.\ \text{degree}((\text{RB/RBR/RBS})\, combination\ with \\ (\text{JJ/RB/RBR/RBS})) \oplus (-(-\text{degree}(not\ /\ never\ /\ ...))^2) \\ if\ \text{degree(JJ/RB/RBR/RBS)>0} \\ eg: not\ very\ good, \text{not extremely high} \\ 2.\ \text{degree}((\text{RB/RBR/RBS})\, combination\ with \\ (\text{JJ/RB/RBR/RBS})) \oplus ((-\text{degree}(not\ /\ never\ /\ ...))^2) \\ if\ \text{degree(JJ/RB/RBR/RBS)<0} \\ eg: not\ very\ bad, \text{not extremely annoyed} \end{cases} \quad (2)$$

$$weights\ of \left[(very\ /\ so\ /\ ...) \begin{bmatrix} combination\ with \\ (\text{RB/RBR/RBS}) \\ combination\ with \\ (\text{JJ/RB/RBR/RBS}) \end{bmatrix} \right]$$

$$= \begin{cases} 1.\ \text{degree}((\text{RB/RBR/RBS})\, combination\ with \\ (\text{JJ/RB/RBR/RBS})) \oplus ((\text{degree}(very\ /\ so\ /\ ...))^2) \\ if\ \text{degree(JJ/RB/RBR/RBS)>0} \\ eg: \text{very}\ very\ good, \text{so extremely high} \\ 2.\ \text{degree}((\text{RB/RBR/RBS})\, combination\ with \\ (\text{JJ/RB/RBR/RBS})) \oplus (-(\text{degree}(very\ /\ so\ /\ ...))^2) \\ if\ \text{degree(JJ/RB/RBR/RBS)<0} \\ eg: \text{very}\ very\ bad, \text{so extremely annoyed} \end{cases} \quad (3)$$

Definition 3.1.3 (Weight for a review). The weight of a review is calculated based on fuzzy operation. The appearance frequency of the opinion features in the review and the related fuzzy weights of opinion words are two important elements that can determine the weight of a review.

$$\text{RW}= \frac{\sum_{i=1}^{n} \text{fuzzy scale}(opinion\ words) \otimes f(\text{ Related features})_i}{\sum_{i=1}^{n} f(\text{Related features})_i} \quad (4)$$

$$(where,\ n\ is\ the\ total\ number\ of\ features\ in\ a\ review)$$

The weights of the extracted opinion expressions are defined in case 1 and case 2, and the weight for a review is defined in definition 3.1.3. Fuzzy logic is used in the calculation process to make sure the obtained weights are accurate. In order to deeply answer the necessary information of an opinion, the opinion words and the features should be accurately extracted. In the next section, the algorithms of opinion words and feature extraction will be given and the dependency structure will be employed to express the relations between opinion expressions and features.

4 Jointly execute opinion mining extraction tasks

4.1 Extraction rules defined based on dependency relations

The extraction is mainly between features and opinion words. For convenience, some symbols are defined easy reusability. The relations: between opinions and features are defined as FO↔Rel, between opinion words themselves are OO↔Rel, and between features are FF↔Rel. Four basic extraction tasks are defined to separate information extraction: (1). Extracting products' features by using opinion words (FO↔Rel); (2). Retrieving opinions by using the obtained features (OF↔Rel); (3). Extracting features by using the extracted features (FF-Rel); (4). Retrieving opinions based on the known opinion words (OO-Rel). Four categories of running rules are clarified and depicted in Table 1.

In Table 1, o (or f) represents for the obtained opinion expressions (or features). O (or F) is the set of known opinions (or features) either given or obtained. POS (O/F) means the POS information that contains the linguistic category of words, such as *noun* and *verb*.{NN, NNS, JJ, RB,VB} are POS tags. O-Dep, that represents the opinion word O, depends on the second word based on O-dep relation. F-dep means the feature word F depends on the second word through F-dep relation. MR={nsubj, mod, prep, obj, conj,dep}, 'mod' contains {amod, advmod}, 'obj' contains {pobj, dobj}. Finally, rules ($R1_i$ –$R4_i$) are formalized and employed to extract features (f) or opinion words (O).

Table 1. Rules for features and opinion expressions extraction

Rule	Input	Representation Formula	Output	Example
$R1_1$	O	$O \xrightarrow{\text{Depend (O-Dep)}} F$; where, $O \in \{O\}$, O-Dep $\in \{MR\}$, $POS(F) \in \{NN, NNS\}$	f=F; FO↔Rel	Canon PowerShot SX510 takes **good** *photos*. (good→amod→photos)
$R1_2$	O	$O \xrightarrow{O-Dep} H \xleftarrow{F-Dep} F$ s.t. $O \in \{O\}$, O / F-Dep $\in \{MR\}$ $POS(F) \in \{NN, NNS\}$	f=F FO↔Rel	The Canon PowerShot SX510 *HS* is a very **good** *value* thanks to a new sensor. (good→amod→value←nsubj←HS)
$R1_3$	O	$O \xrightarrow{O-Dep} H \xrightarrow{F-Dep} F$ s.t. $O \in \{O\}$, O / F-Dep $\in \{MR\}$, $POS(F) \in \{NN, NNS\}$	f=F FO↔Rel	It works **great** for a kindle camera. (great ←prep←for←pobj←camera)
$R2_1$	F	$O \xrightarrow{O-Dep} F$; s.t. $F \in \{F\}$, $POS(O) \in \{JJ, RB, VB\}$	o=O OF↔Rel	Same as $R1_1$, *photos* as the known word and *good* as the extracted word.
$R2_2$	F	$O \xrightarrow{O-Dep} H \xleftarrow{F-Dep} F$ s.t. $f \in \{F\}$, O / F-Dep $\in \{MR\}$ $POS(O) \in \{JJ, RB, VB\}$	o=O OF↔Rel	Same as $R1_2$, *HS* as the known word and *good* as the extracted word, also extract the middle word *value*
$R2_3$	F	$O \xrightarrow{O-Dep} H \xrightarrow{F-Dep} F$ s.t. $f \in \{F\}$, O / F-Dep $\in \{MR\}$ $POS(F) \in \{JJ, RB, VB\}$	o=O OF↔Rel	Same as $R1_3$, *camera* as the known word and *great* as the extracted word. (camera→pobj→for→prep→great)
$R3_1$	F	$F_{i(j)} \xrightarrow{F_{i(j)}\text{-Dep}} F_{j(i)}$ s.t. $F_{j(i)} \in \{F\}, F_{i(j)}\text{-Dep} \in \{conj\}$ $POS(F_{i(j)}) \in \{NN, NNS\}$	f=F FF↔Rel	It takes breathtaking **photos** and great **videos** too. (photos→conj→videos)
$R3_2$	F	$F_{i(j)} \xrightarrow{F_{i(j)}\text{-Dep}} F_{j(i)}$ s.t. $F_{j(i)} \in \{F\}, F_{i(j)}\text{-Dep} \in \{NN\}$ $POS(F_{i(j)}) \in \{NN, NNS\}$	f=F FF↔Rel	The image **quality** is great. quality←nn←image
$R3_3$	F	$F_i \xrightarrow{F_i-Dep} H \xleftarrow{F_j-Dep} F_j$ s.t. $F_i \in \{F\}, F_i / F_j - Dep \in \{MR\}$ $POS(F_j) \in \{NN, NNS\}$	f=F FF↔Rel	**SX500** has a smaller camera and a good sized zoom. (SX500→nsubj→has←dobj←camera←conj←zoom)
$R4_1$	O	$O_{i(j)} \xrightarrow{O_{i(j)}-Dep} O_{j(i)}$, s.t. $O_{j(i)} \in \{O\}$, $O_{i(j)} - Dep \in \{advmod, conj\}$, $POS(O_{i(j)}) \in \{RB\}$	o=O OO↔Rel	Canon PowerShot *SX510* takes significantly **better** indoor *photos*. (better←advmod←significantly) This camera is **light** and easy to hold. (light←conj←easy)
$R4_2$	O	$O_i \xrightarrow{O_i-Dep} H \xleftarrow{O_j-Dep} O_j$, s.t. $O_i \in \{O\}$, $O_i - Dep == O_j - Dep$ $POS(O_{i(j)}) \in \{JJ\}$	o=O OO↔Rel	If anybody wants a new light, smart, easy use camera, I highly recommend Canon PowerShot. (new→amod→camera←amod←light; new→amod→camera←amod←smart;…)

4.2 Opinion mining extraction algorithm

Table 2 shows the detailed opinion mining extraction algorithm. The initial values of the proposed algorithm are shown as: opinions dictionary O, the opinion degree intensifiers OD, and the review data RD. This algorithm adopts a single review from customers as the basic analysis unit. For each review, anytime the customer mentions a feature name, such as camera, those words are considered unique and should be excluded from the analysis. In other words, if the review talks about the “camera’s zoom” feature, and afterwards the same word “zoom” appears again in the same review; the word "zoom" will be excluded from being analyzed further. This assumption determines the stop point of the proposed algorithm. If no new feature words are found in the review, then the algorithm will stop its analysis for the current review and begin to analyze the next review.

Table 2 Algorithm 1: opinion mining extraction algorithm

Algorithm Opinion_Mining_Extraction()

Input: Opinion word dictionary O, Opinion Degree Intensifiers OD, Review Data:RD
Output: The set of features F, the set of expanded opinion words EO, the opinion polarity (or orientation) for a product: OW
BEGIN
Expanded opinion words: EO= $\varnothing$; F= $\varnothing$; ODI= $\varnothing$
For each dependency parsed review RD_k
  for each word tagged JJ,RB, and VB in RD_k
    Traversing the RD_k, and extracting the opinion words (OP_i) if they are appearing in O; i++;
    Extracting new opinion words $\{OP_j\}$ in RD_k by using the Rules $R4_1$-$R4_2$ based on extracted opinion words $\{OP_i\}$; j++;
    Inputting the obtained OP_i and OP_j into EO, and then EO=$\{OP_{[1,\ldots,i]}$, $OP_{[1,\ldots,j]}\}$(for short EO=$\{OP_{1-i}, OP_{1-j}\}$);
    Traversing the RD_k, and extracting the degree intensifier words (DW_d) if they are appearing in OD;
    Inputting the obtained DW_d into ODI, and then ODI=$\{DW_{1-d}\}$; d++;
  End for
  Extracting features $\{F_{fi}\}$ in RD_k by using the Rules $R1_1$-$R1_3$ based on opinion words EO=$\{OP_{1-i}, OP_{1-j}\}$; fi++;
  if (Extracted new features not in F)
    Extracting new features $\{F_{fj}\}$ using Rules $R3_1$-$R3_3$ based on the new extracted features $\{F_{fi}\}$; fj++;
    Extracting and updating new opinion words $\{OP_{1-p}\}$ using Rules $R2_1$-$R2_3$ based on extracted features F=$\{F_{fi}, F_{fj}\}$;
    Extracting new features $\{F_{fp}\}$ in RD_k by using the Rules $R1_1$-$R1_3$ based on new opinion words EO=$\{OP_{1-p}\}$; fp++;
  End if
  Setting F=$\{F_{fi}, F_{fj}, F_{fp}\}$; EO=$\{OP_{1-i}, OP_{1-j}, OP_{1-p}\}$;
  KernelFeature_OpinionSets=Build_kernel(F, EO, RD_k);
Recording appearing frequency af of EO based on related F;
  if the opinion words EO have the corresponding degree intensifier ODI
    Building triple {ODI, EO, F}
  Else if
    Building triple {null, EO, F}
  End if
Unique and update {ODI,EO,F};
Calculating the opinion polarity{OW} based on **Definition 3.1.1- 3.1.3, Triple {ODI, EO, F}, and af;**
End for
END

5 Performance comparison between baseline approaches and proposed method

The similar products S110, SX510 HS, and SX280 HS have 232, 381, and 517 reviewers respectively. The total number of sentences for each dataset is marked in Table 3. For each sentence of each review, it has five rows that include the sentence itself, POS, dependency relations, detailed dependency relations, and the sequence markers. The sequence markers are F, O, D, and N in the data sets. F denotes the features, O denotes the opinion words, DO denotes the opinion degree intensifier words, and N denotes none of them. We generate the experiments results in sentiment classification, feature and opinion extraction to make deeper analysis of algorithms performance. The classification results demonstrate that the proposed method is more effective than the other algorithms. The reason for this is that we clearly defined each opinion words’ fuzzy scale, considering some adverbs and verbs as the opinion words, and finding the modifier that could give the additional intensity information of an opinion word.

In order to test the information extraction performance, we compare the proposed method with Qiu et al. (hereinafter called Qiu2011), and **c**onditional **r**andom **fi**elds (CRF) (Jakob and Gurevych, 2010, hereinafter called Jakob2010).Qiu2011 adopted dependency parser to identify syntactic relations between opinion words and features and proposed a double propagation algorithm to do information extraction. Qiu2011 claimed that the proposed propagation algorithm outperforms CRF significantly (Lafferty et al., 2001), Popescu (Popescu and Etzioni, 2007), and Kanayama (Kanayama and Nasukawa, 2006). Jakob2010 argue that the advanced CRF-based algorithm clearly outperforms the baseline algorithms on all datasets, which improves the performance based on F-score in four single domains. Hence, we employ Qiu2011 and Jakob2010 approach as the baseline.

Table 3 gives the comparison results of different approaches. The precision of feature extraction of our method is 6.43% higher on average than Jakob2011 and Qiu2011 respectively, which means that our method can extract more effective instances among feature elements. The recall of feature extraction is also significantly improved, which is up 29.46% and 14.86% on average by comparing with Jakob2011 and Qiu2011 respectively. We observe our method outperforms better than the other methods for opinion extraction in terms of precision and recall. Meanwhile, the gain in F-score is between 0.6713 in S110 (feature extraction) and 0.8211 in SX510 HS (opinion extraction), and the achieved F-score is higher than the other methods in all datasets. The reason is that we match the reviewed data with an intensive opinion words dictionary and consider the dependency relations up to the phrase level by building kernels between closely related words of each sentence. Although the proposed method clearly outperforms the other baseline approaches, the same generation trend also exists in the individual results: Opinion extraction yields better results than feature extraction. It is because the feature words are more complex and changeable. The opinion words and the obtained feature words are used as guide words to iteratively find new features words, whereas the reviewers may adopt synonyms or analogies to describe the same feature. In general, the comprehensive analysis shows that our method is more effective and more suitable to be used in real-life cases.

Table 3 Precision, Recall, and F-score of our method, Qiu2011, and Jakob2011

Selected Product ID (High-End, Advanced Digital Canon Camera)	Directions	Methods	P	R	F
Canon PowerShot S110 No. of Reviews: 232 Sentences: 2054	Feature extraction	Our method	**0.6575**	**0.6857**	**0.6713**
		Qiu2011	0.6139	0.6043	0.6091
		Jakob2010	0.5714	0.4400	0.4972
	Opinion extraction	Our method	0.7625	**0.8222**	**0.7912**
		Qiu2011	**0.7778**	0.7125	0.7437
		Jakob2010	0.6625	0.7143	0.6874
Canon PowerShot SX510 HS No. of Reviews: 381 Sentences: 2456	Feature extraction	Our method	**0.8046**	**0.6575**	**0.7237**
		Qiu2011	0.7241	0.4118	0.5250
		Jakob2010	0.5172	0.2941	0.3750
	Opinion extraction	Our method	**0.7812**	**0.8654**	**0.8211**
		(Qiu et al., 2011)	0.7677	0.5135	0.6154
		CRF	0.6970	0.4662	0.5587
Canon PowerShot SX280 HS No. of Reviews: 517 Sentences: 4992	Feature extraction	Our method	**0.6892**	**0.7183**	**0.7034**
		Qiu2011	0.6204	0.5986	0.6093
		Jakob2010	0.4599	0.4437	0.4516
	Opinion extraction	Our method	**0.7958**	**0.7434**	**0.7687**
		Qiu2011	0.6069	0.5789	0.5926
		Jakob2010	0.4437	0.4145	0.4286

6 Conclusions and Future Work

In this paper, we proposed an opinion mining extraction algorithm that can jointly identify features, opinion expressions, and feature-opinion by using fuzzy logic to determine opinion boundaries and adopting syntactic parsing to learn and infer propagation rules between opinions and features. Our algorithm allows opinion extraction to be executed at the phrase level and can automatically detect the features that contain more than one word by building kernels through closest words. This work presents opinion intensifier sets that can aid to extract opinion degree words. In addition, we also have discovered more dependency relations between features and opinions than the previous works. Experimental evaluations show that our algorithm outperforms the baseline approaches on different extraction tasks. Recognition of important features based on the proposed algorithm will be further studied. Meanwhile, identification of proper features to improve for both product orientation and consumption quantities will be analyzed deeper in the future work as well.

7 References

[1] Bethard, S., Yu, H., Thornton, A., Hatzivassiloglou, V., Jurafsky, D., 2006. Extracting opinion propositions and opinion holders using syntactic and lexical cues, in: Computing Attitude and Affect in Text: Theory and Applications. Springer, pp. 125–141.

[2] Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.

[3] Brody, S., Elhadad, N., 2010. An unsupervised aspect-sentiment model for online reviews, in: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, pp. 804–812.
[4] Choi, Y., Cardie, C., 2010. Hierarchical sequential learning for extracting opinions and their attributes, in: Proceedings of the ACL 2010 Conference Short Papers. Association for Computational Linguistics, pp.269–274.
[5] Hofmann, T., 1999. Probabilistic latent semantic indexing, in: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, pp.50–57.
[6] Hu, Mingqing and Bing Liu. 2004. Mining and summarizing customer reviews. In Proceedings of IGKDD'04, pages168–177.
[7] Jakob, N., Gurevych, I., 2010. Extracting Opinion Targets in a Single- and Cross-domain Setting with Conditional Random Fields, in: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP '10. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 1035–1045.
[8] Jin, W., Ho, H.H., 2009. A novel lexicalized HMM-based learning framework for web opinion mining, in: Proceedings of the 26th Annual International Conference on Machine Learning. Citeseer, pp. 465–472.
[9] Joshi, M., Penstein-Rosé, C., 2009. Generalizing dependency features for opinion mining, in: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers. Association for Computational Linguistics, pp. 313–316.
[10] Kanayama, H., Nasukawa, T., 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis, in: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp. 355–363.
[11] Kobayashi, N., Inui, K., Matsumoto, Y., 2007. Extracting Aspect-Evaluation and Aspect-Of Relations in Opinion Mining., in: EMNLP-CoNLL.pp.1065–1074.
[12] Lafferty, J., McCallum, A., Pereira, F.C., 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
[13] Okanohara, D., Miyao, Y., Tsuruoka, Y., Tsujii, J., 2006. Improving the scalability of semi-markov conditional random fields for named entity recognition, in: Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, pp. 465–472.
[14] Popescu, A.-M., Etzioni, O., 2007.Extracting product features and opinions from reviews, in: Natural Language Processing and Text Mining.Springer,pp.9–28.
[15] Qiu, G., Liu, B., Bu, J., Chen, C., 2011. Opinion word expansion and target extraction through double propagation. Comput. Linguist. 37, 9–27.
[16] Titov, I., McDonald, R.T., 2008. A Joint Model of Text and Aspect Ratings for Sentiment Summarization., in: ACL. Citeseer, pp. 308–316.
[17] Wu, Y., Zhang, Q., Huang, X., Wu, L., 2009. Phrase dependency parsing for opinion mining, in: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3. Association for Computational Linguistics, pp. 1533–1541.
[18] Yang, B., Cardie, C., 2012.Extracting opinion expressions with semi-markov conditional random fields, in: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational Linguistics,pp.1335–1345.
[19] Zhang, H., Sekhari, A., Ouzrout, Y., Bouras, A., 2014. Optimal Inconsistency Repairing of Pairwise Comparison Matrices Using Integrated Linear Programming and Eigenvector Methods. Math. Probl. Eng. 2014, e989726. doi:10.1155/2014/989726
[20] Zhuang, L., Jing, F., Zhu, X.-Y., 2006. Movie review mining and summarization, in: Proceedings of the 15th ACM International Conference on Information and Knowledge Management. ACM, pp. 43–50.