Survey on Various methods of Opinion Target Extraction

[#1]Ms. Sunita Patil, [#2]Mrs. N. S. Patil

[1]sunitajpatil92@gmail.com
[2]snilampatil2012@gmail.com

[#12]Department of Computer Engineering

D. Y. Patil College, Akurdi

ABSTRACT

Opinion mining also called as sentiment analysis, has enticed a great deal of attention recently due to many practical applications and challenging research problems. Opinion mining can be useful in several ways like it can help marketers to evaluate the success of particular product or services provided by particular resource or particular features of the product, also used for deciding which release of product or service are most popular and also used for determining which features are liked or disliked by the customer. Extracting opinion targets is one of the important problems in opinion mining, the targets are the objects on which users have expressed their opinions, usually as nouns or noun phrases. This task is very important because customers are usually unsatisfied with only the overall sentiment polarity of a product, but expect to find the fine-grained sentiments about an aspect or a product feature mentioned in reviews. There are different approaches to overcome from this problem. In this paper, we reviewed a several ways to accomplish this and presented a novel approach to solve this problem.

Index Terms: Double propagation, Opinion mining, Opinion target Extraction, PSWAM, WTM

ARTICLE INFO

I. INTRODUCTION

Data Mining is defined as the process of extracting information from large sets of data. Text mining is one application of data mining. Text mining is the process of deriving high-quality information from text. To get valuable structured information, text data mining can be search carefully and systematically through a text document or resource. There are mainly two categories of textual information one is facts and another is opinions. Facts are objective statements about entities and events in the world. Opinions are subjective statements that reflect peoples sentiments or perceptions about the entities and events. Opinion mining is a type of natural language processing for tracking the mood of the public about a particular product. Opinion mining involves building a system to collect and categorize opinions about a product.

A very basic step of opinion mining is feature extraction. Fig. 1 shows the process of opinion mining. In opinion mining, one fundamental problem is opinion target extraction. Target extraction is nothing but to extract items on which opinion are expressed. Opinion targets are usually nouns/noun phrases. In online product reviews, opinion targets are products or product features. So opinion target extraction is also called as feature extraction. For example, in the opinion sentence I am not happy with the battery life of this phone, battery life is the target of the opinion. The feature
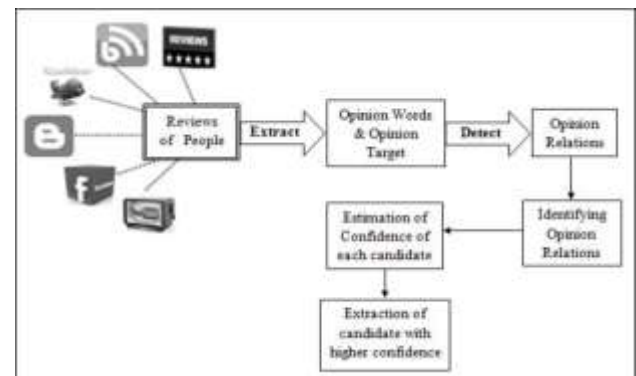

Fig. 1. System Architecture

extraction is important because if we dont know the targets, it is not so much useful of opinions expressed in a sentence

or document. So opinion target extraction is very important task. The feature extraction is one of main challenge in opinion mining.

In opinion mining there are several challenges such as object identification, feature extraction, grouping synonyms, opinion orientation classification, integration, opinion mining at document level. Among them feature extraction is problem in which considering the sentence The voice on my phone was not clear the object feature is voice. In this technique, frequently used nouns noun phrases as features are identified as features, which are usually genuine features. Opinion targets usually are product features or attributes. Accordingly this subtask is also called as product feature extraction [3], [4].

The rest of the paper structure is as follows. In section II, related work is given. Section III, we briefly describe the details of opinion mining.

II. REVIEW OF LITERATURE

In opinion mining, opinion target extraction is crucial task. Many methods have been studied for the product review mining. Hu and Liu [14] are consider the frequent nouns and noun phrases as product feature candidates. In [15], for product feature extraction, authors proposed a language model. In this model he assumed that product features are mentioned more often in a product review than they are mentioned in general English text. However, statistics may not be reliable

when the corpus is small, as pointed out earlier. The work by Kobayashi, Inui, and Matsumoto [16] focused on the aspect-evaluation and aspect-of extraction problems in blogs. Double propagation method is differs from theirs in that they make use of syntactic relations from dependency trees. The author consider relations of opinion targets and opinion words as well as many types of relations. In the work of Stoyanov and Cardie [17], the authors treated target extraction as a topic coreference resolution problem. To cluster opinions sharing the same target together is the essential to this method. They proposed to train a classifier to judge if two opinions are on the same target, which indicates that their approach is supervised. Double propagation method is semi-supervised so it differs from other method. Opinion target extraction have method such as [18], [19] ; Liu et al., 2006; Popescu et al., 2005; Wu et al., 2005; Wang et al., 2008; Li et al., 2010; Su et al., 2008; Li et al., 2006). In general, the existing approaches can be divided into two main categories: supervised and unsupervised methods. In supervised approaches, the opinion target extraction task was usually regarded as a sequence labeling task (Jin et al. 2009; Li et al. 2010; Wu et al., 2009; Ma et al. 2010; Zhang et al., 2009).

III. DETAILS OF OPINION MINING

A. Overview of Opinion mining

The valuable things of the marketing places is to identify and subsequently satisfy consumer needs. So, to examine con-sumer needs and to implement effective

marketing strategies aimed at satisfying these needs, marketing managers need relevant, current information about consumers, competitors and other forces in the marketplace. There is not much study on opinions in the past, as a significant part of consumer information, which is present in companys own databases, has been ignored in the past and there was less opinionated text available before the existence of World Wide Web. Nowadays users of web 2.0 contribute content actively in web-forums and product review websites. With the growth of the web over the last decade, opinions can now be found almost everywhere-blogs, social networking sites like Facebook and Twitter, news portals, e-commerce sites, etc. Therefore, to get additional market insights, modern companies have a strong need to utilize this user-generated content. So, if a person wants to purchase a certain product or companies want to know opinions of the consumers about their products, web reviews can be used for this purpose as the web has acquired immense value as an actively evolving repository of knowledge for market research. [1] [2]. Due to availability of large volume of information on Web, opinion mining can still be a formidable task.

There are many applications opinion mining such as pur-chasing product or service. While purchasing a product or service, taking right decision is no longer a difficult task. By this technique, people can easily evaluate others opinion and experience about any product or service and also he can easily compare the competing brands. Now people dont want to rely on external consultant. The Opinion mining and sentiment analysis extract people opinion form the huge collection of unstructured content, the internet, and analyze it and then present to them in highly structured and under-standable manner. Another application of opinion mining is quality improvement in product or service. By Opinion mining and sentiment analysis the manufactures can collect the critics opinion as well as the favorable opinion about their product or service and thereby they can improve the quality of their product or service. They can make use of online product reviews from websites such as Amazon and C—Net [6, 7], RottenTomatoes.com [8] and IMDb [9]. Marketing research is also one of the application of opinion mining. The result of sentiment analysis techniques can be utilized in marketing research [6]. By sentiment analysis techniques, the recent trend of consumers about some product or services can be analyzed. Similarly the recent attitude of general public towards some new government policy can also be easily analyzed. These all result can be contributed to collective intelligent research.

B. Problems in Opinion mining

Since opinion mining is a relatively new filed, thus there are several challenges to be faced. According to Reference [4] current techniques are just primitive for opinions and com-parisons identification and extraction. Mainly these challenges are related to the authenticity of the extracted data and the methods used in it. Some challenges of opinion mining are as follows:
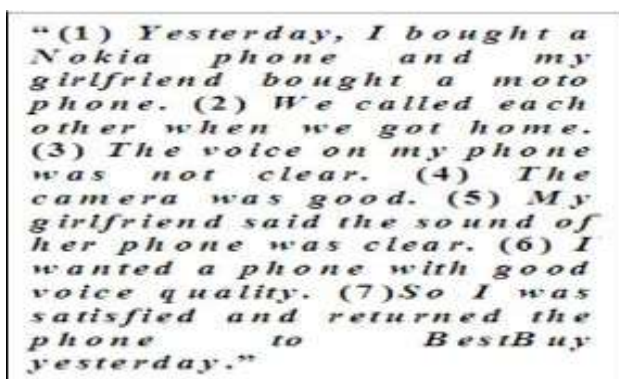
Object identification: In opinion mining, firstly you have to identify the objects in a review on which opinion

have been given. This problem is important because without knowing the object on which an opinion has been expressed, the opinion is of little use. However, there is a difference. In for opinion mining, only those objects in the review are to be considered which are in competition to each other. The system thus needs to separate relevant objects and irrelevant objects.

Feature extraction: In the review shown in Figure 9, considering the sentence The voice on my phone was not clear the object feature is voice. Reference [10] discusses such a noun based approach in which supervised pattern mining method is suggested. In this technique, frequently used nouns noun phrases as features are identified as features, which are usually genuine features. Many other techniques are also used for extracting information, e.g., conditional random fields (CRF), hidden Markov models (HMM), and many others.

Grouping synonyms: Different words or phrases can be used to refer to the same feature of the object. So, such words (synonyms) should be identified and grouped together. It is a difficult task to identify these words. A lot of research is required to be done on this issue as it has not been much addressed in the past. To produce a summary similar to the one in Fig. 2, it is needed to group synonym features, as people often use different words or phrases to describe the same feature. In this example, voice and sound both refer to the same feature.

Opinion orientation classification: This task identifies the orientation of opinions i.e. determines whether the opinions on the features are positive, negative or neutral. In the review shown in Fig. 2, the opinion on voice is negative. Many approaches can be used for this purpose.

> "(1) Yesterday, I bought a Nokia phone and my girlfriend bought a moto phone. (2) We called each other when we got home. (3) The voice on my phone was not clear. (4) The camera was good. (5) My girlfriend said the sound of her phone was clear. (6) I wanted a phone with good voice quality. (7) So I was satisfied and returned the phone to BestBuy yesterday."

Fig. 2. A complex example blog.

Usually, lexicon based approach is used as it performs quite well. The lexicon-based approach basically uses opinion words and phrases in a sentence to determine the orientation of an opinion on a feature. A relaxation labeling based approach is also proposed. To solve the problem of feature extraction known as opinion target ex-traction, three approaches are double propagation, word-based translation model(WTM) and partially-supervised word alignment model (PSWAM).

IV. TECHNIQUES

For opinion target extraction three methods are describe in this section.

A. Double Propagation

The basic idea of double propagation is to extract opinion words (or targets) iteratively using known and extracted (in previous iterations) opinion words and targets through the identification of syntactic relations. The identification of the relations is the key to the extractions. As this approach propagates information back and fort between opinion words and targets, author call it double propagation. Target extraction tasks perform iteratively base on propagation using the rela-tions defined herein. To bootstrap the propagation, it require a seed opinion lexicon. Extraction approach adopts the rule-based strategy which is quite natural given the well-defined relations. For example, in an opinion sentence Canon G3 takes great pictures, the adjective great is parsed as directly depend-ing on the noun pictures through mod, formulated as a OT-Rel quadruple DD, mod. If we know great is an opinion word and are given a rule like a noun on which an opinion word directly depends through mod is taken as the target, can easily extract pictures as the target. Similarly, if we know pictures is a target, we could extract the adjective great as an opinion word using a similar rule. Based on such observations, the idea of the whole propagation approach is first to extract opinion words and targets using the seed opinion lexicon and then use the newly extracted opinion words and targets for further target and opinion word extraction. The propagation ends until no more new opinion words or targets can be identified. In this way, even if the seed opinion lexicon is small, targets can still be extracted with high recall. The extraction is performed using identified relations between opinion words and targets, and also opinion words/targets themselves. The relations are described syntactically based on the dependency grammar. The advantage of the double propagation method is that it only needs an initial opinion lexicon to start the bootstrapping process. Thus, the method is semi-supervised due to the use of opinion word seeds. But there is some disadvantage of this method. Such that they exploited direct dependency relations between words to extract opinion targets and opinion words iteratively. The main limitation of Qius method is that the patterns based on dependency parsing tree may introduce many noises for the large corpora. This limitation can be overcome by the WTM.

B. WTM

Word-based translation model (WTM) can overcome the drawback of double propagation method. In this method first it identifying opinion relations between opinion targets and opinion words as a word alignment task. Author argue that an opinion target can find its corresponding modifier through monolingual word alignment. For example in Fig. 3, the opinion words colorful and amazing are aligned with the target screen through word alignment. WTM is used to
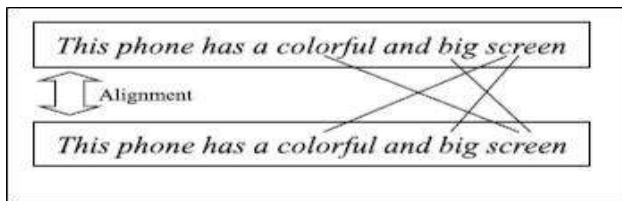
Fig. 3. Mining opinion relations between words using the word alignment model [1].

perform monolingual word alignment for mining associations between opinion targets and opinion words. In this process, several factors, such as word co-occurrence frequencies, word positions etc., can be considered globally. Compared with adjacent methods, [11] [12] [13] WTM doesnt identify opinion relations between words in a given window, so long-span relations can be effectively captured. Compared with syntax-based methods, without using parsing, WTM can effectively avoid errors from parsing informal texts. So it will be more robust. In addition, by using WTM, this method can capture the one-to-many or many-to-one relations (one-to-many means that, in a sentence one opinion word modifies several opinion targets, and many-to-one means several opinion words modify one opinion target). Thus, its reasonable to expect that WTM is likely to yield better performance than traditional methods for mining associations between opinion targets and opinion words. Based on the mined associations, extract opinion targets in a ranking framework. All nouns/noun phrases are regarded as opinion target candidates. Then a graph-based algorithm is exploited to assign confidences to each candidate, in which candidate opinion relevance and importance are incorporated to generate a global measure. At last, the candidates with higher ranks are extracted as opinion targets. Compared with most traditional methods (Hu et al. 2004; Liu et al., 2005; Qiu et al., 2011), thi method can not extract opinion targets iteratively based on the bootstrapping strategy, such as Double Propagation, instead all candidates are dynamically ranked in a global process. Therefore, error propagation can be effectively avoided and the performance can be improved.

A graph based approach is used to extract opinion targets using WTM. So that opinion targets are extracted in a global process, which can effectively solve the problem of error propagation which is not overcome by the double propagation. As compared to previous methods such as adjacent methods and syntax-based methods, WTM method can capture opinion relations more precisely and therefore be more effective for opinion target extraction, especially for large informal Web corpora. But WTM method is completely unsupervised man-ner, which makes the alignment quality still unsatisfactory. So to solve this problem PSWAM approach is used.

C. PSWAM

PSWAM is one of the best method for opinion target extraction. For example in Fig. 4, colorful is incorrectly regarded as the modifier for phone if the WAM is performed in an whole unsupervised manner ((a) in Fig 4).

However, it can assert colorful should be aligned to screen ((b) in Fig. 4). Then the errors in (a) can be corrected by using this partial links to supervise the statistical model, as shown in (c) in Fig. 5 Thus, this approach is to extract opinion targets by using
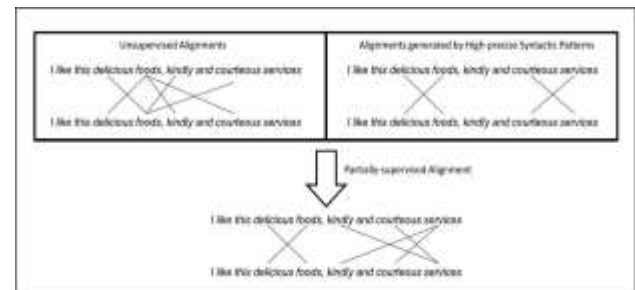

Fig. 4. Mining opinion relations between words using the word alignment model. [1]

partially-supervised word alignment model (PSWAM). First use some high-precise-low-recall syntactic patterns to capture partial opinion relations (partial alignment links) in sentences. Although existing syntactic parsing algorithms cannot obtain the precise whole syntactic tree of the informal sentences, they believe some short or direct dependency relations between words can be still obtained precisely. Then these extracted partial alignment links would be regarded as ground truths. And an constrained EM algorithm based on hill-climbing is performed to determine all alignments in sentences, where the model will be consistent with these links as far as possible. In this way, more correct opinion relations can be mined. This model can not only inherit the advantages of word alignment model: considering multiple factors (word co-occurrence frequencies, word positions etc.) in the global process, effectively avoiding noises from syntactic parsing errors when dealing with informal texts, but also can improve the mining performance by using partial supervision. Thus, its reasonable to expect that PSWAM is likely to yield better performance than traditional methods. Then extract opinion targets in a graph-based framework based on the mined associations. All nouns (noun phrases) are regarded as opinion target candidates. A bipartite graph is constructed to model the opinion relations between words. Assume that two candidates are modified by similar opinion words, they are likely to belong to the similar category. If have we known one of them is an opinion target, the other one has high probability to be an opinion target. Thus, the opinion target confidence can propagate among vertices. A random walk algorithm can be applied to estimate the confidence of each candidate, and the candidates with higher confidence will be extracted as the opinion targets. However, in traditional random work algorithm, it observe that the higher-degree vertices are prone to collect more information from other vertices and put more impacts on other vertices. These words usually are general words and may introduce noises. For example, the opinion word good, may be used to modify multiple objects like good design, good feeling and good things. The degree of good will be high in the graph. If we have know that the design has higher confidence to be an opinion target, its confidence will be propagated to feeling and thing through good. As a result, feeling and thing will

probably to be given higher confidence as opinion targets. Its unreasonable. To resolve this problem, author make penalty on the higher-degree vertices to weaken the impacts of them and decrease the probability of the random walk running into the unrelated regions in the graph. In this way, errors can be effectively avoided.

V. CONCLUSION

This survey paper explains various techniques available for opinion target extraction. Each of these techniques has their own advantages and disadvantages. But all these techniques can check the relation of words in the sentences. So for topical relation it is required to developed the technique.

ACKNOWLEDGMENT

REFERENCES

[1] Liu, Kang, Liheng Xu, and Jun Zhao. "Co-extracting opinion targets and opinion words from online reviews based on the word alignment model." IEEE Transactions on knowledge and data engineering 27.3 (2015): 636-650.

[2] Liu, Bing, and Lei Zhang. "A survey of opinion mining and sentiment analysis." Mining text data. Springer US, 2012. 415-463.

[3] Hu, Minqing, and Bing Liu. "Mining opinion features in customer reviews." AAAI. Vol. 4. No. 4. 2004.

[4] Popescu, Ana-Maria, and Orena Etzioni. "Extracting product features and opinions from reviews." Natural language processing and text mining. Springer London, 2007. 9-28.

[5] M. Hu and B. Liu, Mining opinion features in customer reviews, in Proc. 19th Nat. Conf. Artif. Intell., San Jose, CA, USA, 2004, pp. 755760.

[6] 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, 2008, pp. 289295.

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

[8] K. Liu, L. Xu, and J. Zhao, Opinion target extraction using wordbased translation model, in Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn., Jeju, Korea, Jul. 2012, pp. 13461356.

[9] R. C. Moore, A discriminative framework for bilingual word alignment, in Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process., Vancouver, BC, Canada, 2005, pp. 8188.

[10] 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, 2010, pp. 653661.

[11] 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, 2009, pp. 465472.

[12] T. Ma and X. Wan, Opinion target extraction in chinese news comments. in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 782790.

[13] F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu, Cross-domain coextraction of sentiment and topic lexicons, in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Jeju, Korea, 2012, pp. 410419.

[14] 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, 2005, pp. 339346.

[15] 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, 2009, pp. 11991204.

[16] 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. 927, 2011.

[17] L. Zhang, B. Liu, S. H. Lim, and E. OBrien-Strain, Extracting and ranking product features in opinion documents, in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 14621470.

[18] J. M. Kleinberg, Authoritative sources in a hyperlinked environment, J. ACM, vol. 46, no. 5, pp. 604632, Sep. 1999.

[19] Brown, Peter F., et al. "The mathematics of statistical machine transla-tion: Parameter estimation." Computational linguistics 19.2 (1993): 263-311.