

A Survey on Opinion Mining for Extraction of Opinion Word and Opinion Target

D. Menaka¹, E. Saravana Kumar²

P.G Scholar, Department of Computer Science & Engineering, Adhiyamaan College of Engineering, Hosur, India¹

Associate Professor, Dept. of Computer Science & Engineering, Adhiyamaan College of Engineering, Hosur, India²

Abstract: Data mining - an analytical process designed to explore data in which the opinion mining deals with the computational treatment of opinion, sentiment and subjective in text. The main application of opinion mining is collecting the online reviews about the product, social networks informal text. The research problem is extracting the opinion targets and the opinion words and detecting the opinion relations among the words. A novel approach based on the partially supervised alignment model for identifying the opinion relations as an alignment process have been proposed to satisfy the long span relations. To precisely mine the opinion relations among words, the Word Alignment Model (WAM) is used and to progress the error propagation, the graph based co-ranking algorithm is motivated. By comparing with the syntax based method, the word alignment model effectively reduces the parsing errors and the co-ranking algorithm decreases the error probability. The datasets CRD, COAE 2008 and Large are used in various methods. The survey shows the algorithm effectively outperforms when compare to previous methods.

Keywords: Data mining, Opinion mining, WAM, Opinion word, Opinion target.

I. INTRODUCTION

Data mining is the process of collecting, searching through, and analysing a large amount of data in a database, as to discover patterns or relationships. A series of challenges have emerged in data mining and in that one of the major challenges is opinion mining. Opinion mining is the field of study that analyses the people opinions, sentiments, appraisals and emotion towards the entities such as products, services.

The main objective is to gathering the opinion about the products from the online review websites. The emergence of user-generated content via social media had an undeniable impact on the commercial environment. In fact, social media has shifted the content publishing from business towards the customer.

With the explosive growth of social media for like micro-blogs, amazon, flipkart. On the web, individuals and organizations are increasingly using the content in these media for decision making. Each site typically contains a huge volume of opinion text.

The average human reader will have difficulty in identifying the relevant sites and extracting and summarizing the opinions in them. So automated sentiment analysis systems are needed.

In general, sentiment analysis has been classified at three levels. First level is document level, classifies whether a whole opinion document expresses a positive or negative opinion about the product. Second level is sentence level, classifies whether each sentence express a positive, negative or neutral opinion. Third level is aspect level, performs a fine grained classification of an opinion about the product.

In opinion mining, the fundamental subtasks are extracting the opinion word opinion target. Opinion target is a noun

or noun phrases defined as the object about which user express their opinions. Opinion word is a verb or adjectives used to express users' opinion about the object.

For example:

“This phone has an amazing and big screen”

Here, the customers are expect to know whether this review express the positive opinion or negative opinion about the phone. To achieve this aim, the extraction of opinion word and opinion target should be detected. After that, an opinion target list and an opinion word list should be extracted. In above example, the “screen” is the opinion target and the “amazing”, “big” are opinion words for that particular review [1].

After the extraction, the next step is to provide the relation among those words [1]. For this process, the graph co-ranking algorithm [13] is used and the opinion relation graph is constructed to provide the relations among them.

This survey paper is arranged as follows. In section 2, the related work is described .then the algorithms are discussed in section 3 and 4. In section 5, datasets and evaluation metrics are described. The conclusion is described finally in section 6.

II. RELATED WORK

In related work, the extraction of opinion word and opinion target is the old process in opinion mining. This extraction has been widely focused on several techniques as follows and refer table 1.

M. Hu and B. Liu (2007) have proposed a sentiment based classification. The main objective is identifying the opinion sentence from reviews and deciding whether each opinion sentence is positive or negative and summarizing the results [2]. This method extracts the opinion sentences from review.

Year	Publication title	Methodology	Metrics considered
2007	Mining and summarizing customer reviews	Sentiment based classification	Precision and recall.
2012	Cross-domain co extraction of sentiment and topic lexicons	Relational Adaptive bootstrapping (RAP) algorithm	Precision and recall.
2010	Extracting and Ranking Product Features in Opinion Documents	Syntax-based method.	Precision
2012	Opinion Target Extraction Using Word-Based Translation Model	Word based translation model (WTM).	Precision, recall and F-measure
2008	Mining Opinion Features In Customer Reviews	Nearest neighbor rules.	Precision and recall.
2011	Opinion Word Expansion and Target Extraction through Double Propagation	Syntax-based method.	Precision recall, f- Score.
2011	A Simple Word Trigger Method for Social Tag Suggestion	Word trigger method (WTM)	Precision Recall F-measure
2011	A Semi-supervised Word Alignment Algorithm with Partial Manual Alignments	Constrained hill-climbing algorithm	Precision Recall
2010	Opinion target extraction in Chinese news comments	Focused Concepts model	Precision
2008	Bootstrapping Both Product Features and Opinion Words from Chinese Customer Reviews with Cross-Inducing	Iterative Learning Method.	Precision recall, f-measure.
2010	Structure-Aware Review Mining and Summarization	Structure Aware Model Conditional Random Fields	Precision, recall and f-measure.
2013	Opinion target extraction using partially-supervised word alignment model	Word Alignment model	Precision, recall and f-measure.
2007	Extracting Product Features and Opinions from Reviews	Word Semantic Orientation	Precision, recall.

TABLE 1: LIST OF OPINION WORD AND OPINION TARGET EXTRACTION BASED PAPERS

F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu (2012) have proposed a Relational Adaptive bootstrapping (RAP) algorithm [3]. The objective is extracting the sentiment word from the text and generating the seed. This model precisely generates only the seed word (opinion target).

L. Zhang, B. Liu and S. H. Lim (2010) have proposed the Syntax based method to capturing the relation and ranking the product [4]. This method is effectively provides the relations among words for formal text.

K. Liu, L. Xu, and J. Zhao (2012) have proposed the Word based translation model (WTM). The main objective is extracting opinion targets in document level from the reviews [5]. This method is precisely mine only the opinion targets.

Z. Liu, X. Chen, and M. Sun (2011) have introduced a Word trigger method (WTM) to suggest tags according to the text description of a resource [6]. By considering both the description and tags of a given resource as summaries. This method provides the WTM model for summarizing the tags and description of the text.

Q. Gao, N. Bach, and S. Vogel (2011) have proposed a constrained hill-climbing algorithm [7].

The main objective is extracting the opinion targets and providing high precision and low recall. They used precision and recall is used as an evaluation metrics.

B. Wang and H. Wang (2008) have proposed an Iterative Learning Method [8]. The task of identifying product features with opinion words and learning opinion words through features alternately and iteratively. This model extracts only the opinion words.

T. Ma and X. Wan (2010) have used a Focused Concepts model. The main purpose is extracting explicit and implicit opinion targets from news comments [9]. It extracts the implicit and explicit opinion targets.

F. Li, C. Han, M. Huang, X. Zhu, Y. Xia, S. Zhang, and H. Yu (2010) have described a Structure Aware Model Conditional Random Fields [10]. The process of summarizing the review based on document level extraction and extracts positive opinions, negative opinions and object features for review sentences. This model based on document level extraction.

A.-M. Popescu and O. Etzioni (2007) have proposed a Word Semantic Orientation [11]. The main objective is identifying product features and determines the polarity of opinions. The datasets CRD and Large are used.

Even though, several methods are proposed for the extraction of opinion word and opinion target from online reviews have some problems. In order to improve the precision and recall evaluation metric, the Word alignment Model (WAM) and Graph Co-Ranking algorithms are suggested with some other features.

III. WORD ALIGNMENT MODEL (WAM)

WAM method is based on the monolingual model, which precisely mine the opinion relations among the words.

“This phone has an amazing and colourful screen”

Based on WAM, the opinion word and opinion target was extracted. In the above example, “amazing” and “colourful” is the opinion target and the “screen” is an opinion word [1]. When compare to previous method syntactic patterns [3], the WAM precisely mine the words and target.

The previous nearest-neighbour [5] method precisely mines the relation for short span sentences. But WAM method precisely mine relation for both short span and long span relations. The WAM method has some following constrains [1]:

- Nouns/noun phrases should be aligned with adjectives/verbs/a null word.
- Other unrelated words, such as prepositions conjunctions and adverbs should be aligned only with themselves.

Then the hill-climbing algorithm is used to perform local optimizations. For calculating the associations among the words are estimated by

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

Where, w_t means the opinion target and w_o means the opinion word, and then $P(w_t | w_o)$ means the problem between these two words. The above formula was referred from [1].

IV. GRAPH CO-RANKING ALGORITHM

After extracting the opinion word and the opinion target, the relations has been constructed by the opinion relation graph [1] was shown in fig 1.

Graph co-ranking method is estimated by candidate confidence of each opinion word and opinion target and this can be constructed on the graph. The word which has higher problem will be extracted as opinion word or opinion target.

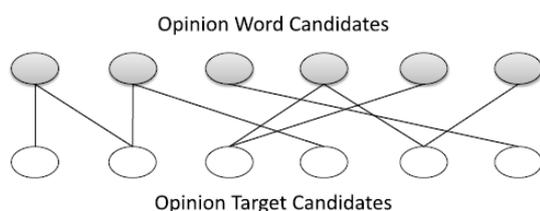


Fig 1. Opinion Relation Graph

The candidate confidence can be estimated by random walking method. Here the confidence of an opinion target candidates and opinion word candidates in the iterations, then the higher confidence than the threshold are obtained as an opinion word or opinion target. The previous bootstrapping method has the error propagation problem. The graph based co-ranking algorithm effectively decreases the error problem [1].

The following features are used to represent the candidates [1]:

- *Saliency feature*: This feature indicates the saliency degree of the candidates.
- *Domain relevance feature*: The opinion targets are domain specific and the difference between them has different domains.

V. DATA SETS AND EVALUATION METRICS

The three datasets are selected to evaluate the WAM method. The datasets are CRD, COAE, and Large [1]. The first customer review data (CRD) has the reviews for five products and the second dataset COFE 2008 contains the Chinese reviews for four products include camera, car, laptop, and phone [3], [4]. The last dataset large has the reviews on three domains includes restaurant, hotel, mp3 [11], [15].

Three annotators are used in the annotation process for proposed methods. The two annotators were involved to extract whether noun/noun phrase is an opinion target or not. Here the conflict may occur. Then the third annotator is used to extract the final results.

The previous methods like nearest-neighbour [5], syntactic pattern [3], double propagation, Word Translation Model [4] were also used in the same datasets.

The three evaluation metrics are selected. The metrics are precision (P), recall (R) and F-measure (F) [1], [8].

Precision is the percentage of selected items that are correct and the recall is the percentage of the correct items that are selected. A combined measure that assesses the precision and recall trade-off is the F-measure.

VI. CONCLUSION

Due to the high usage of internet, the extraction of huge volume of reviews about a product from the online websites to clarify the users taught is increasing day by day. To overcome this problem, the extraction of words and targets and providing relation among these words were followed. These processes has implemented by WAM and Graph based Co-Ranking algorithm and achieves the higher precision when compare to previous methods.

REFERENCES

- [1] K. Liu, L. Xu, and J. Zhao, “Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model,” IEEE Transactions on knowledge and data engineering, 2015, vol. 27, no. 3.
- [2] M. Hu and B. Liu, “Mining and summarizing customer reviews,” in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Seattle, WA, USA, 2004, pp. 168–177.
- [3] F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu, “Cross-domain co extraction of sentiment and topic lexicons,” in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Jeju, Korea, 2012, pp. 410–419.
- [4] L. Zhang, B. Liu, S. H. Lim, and E. O’Brien-Strain, “Extracting and ranking product features in opinion documents,” in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 1462–1470.
- [5] K. Liu, L. Xu, and J. Zhao, “Opinion target extraction using word based translation model,” in Proc. Joint Conf. Empirical Methods

- Natural Lang. Process. Comput. Natural Lang. Learn., Jeju, Korea, Jul. 2012, pp. 1346–1356.
- [6] M. Hu and B. Liu, “Mining opinion features in customer reviews,” in Proc. 19th Nat. Conf. Artif. Intell., San Jose, CA, USA, 2004, pp. 755–760.
- [7] G. Qiu, L. Bing, J. Bu, and C. Chen, “Opinion word expansion and target extraction through double propagation,” *Comput. Linguistics*, vol. 37, no. 1, pp. 9–27, 2011.
- [8] 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. 289–295.
- [9] 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. 653–661.
- [10] T. Ma and X. Wan, “Opinion target extraction in Chinese news comments,” in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 782–790.
- [11] W. X. Zhao, J. Jiang, H. Yan, and X. Li, “Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid,” in Proc. Conf. Empirical Methods Natural Lang. Process., Cambridge, MA, USA, 2010, pp. 56–65.
- [12] Z. Liu, X. Chen, and M. Sun, “A simple word trigger method for social tag suggestion,” in Proc. Conf. Empirical Methods Natural Lang. Process., Edinburgh, U.K., 2011, pp. 1577–1588.
- [13] Q. Gao, N. Bach, and S. Vogel, “A semi-supervised word alignment algorithm with partial manual alignments,” in Proc. Joint Fifth Workshop Statist. Mach. Translation MetricsMATR, Uppsala, Sweden, Jul. 2010, pp. 1–10.
- [14] K. Liu, H. L. Xu, Y. Liu, and J. Zhao, “Opinion target extraction using partially-supervised word alignment model,” in Proc. 23rd Int. Joint Conf. Artif. Intell., Beijing, China, 2013, pp. 2134–2140.
- [15] Z. Hai, K. Chang, J.-J. Kim, and C. C. Yang, “Identifying features in opinion mining via intrinsic and extrinsic domain relevance,” *IEEE Trans. Knowledge Data Eng.*, vol. 26, no. 3, p. 623–634, 2014.
- [16] J. Zhu, H. Wang, B. K. Tsou, and M. Zhu, “Multi-aspect opinion polling from textual reviews,” in Proc. 18th ACM Conf. Inf. Knowl. Manage., Hong Kong, 2009, pp. 1799–1802.