ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

Opinion Word and Opinion Target Extraction using Semi-Supervised Word Alignment Model

Soha N.Chaudhari [1], Prof. Manisha M. Naoghare[2]

M.E. Student, Dept. of Computer Engineering, SVIT, Chincholi, Nashik, Maharashtra, India[1]

Assistant Professor, Dept. of Computer Engineering, SVIT, Chincholi, Nashik, Maharashtra, India[2]

ABSTRACT: In today's era of web, most of the information is expressed in the form opinions. The opinions are widely expressed in various ways like reviews of customer on products. The main problem is determining the semantically orientations of opinions expressed on product features in reviews, which needs to be studied. Hence the important task of consists of detecting opinion relation among words is studied here. Therefore a semi-supervised word alignment is proposed. The opinion word and target is regarded as a candidate, then a confidence is assigned and candidates with higher confidence greater than a threshold are abstracted as the opinion targets or opinion words. The confidence of each candidate is determined collectively by its influencing neighbours, depending upon on the opinion associations between them. The process is modelled by a graph called Opinion Relation Graph. Due to this abstraction, the opinion words and opinion targets which are detected provide an easy assessment to the users.

KEYWORDS: Data mining; opinion words; word alignment model; opinion mining

I. Introduction

There is a large expansion of e-commerce, more and more products are sold on the Web, and more and more people are also buying products online. In order to enhance customer satisfaction and shopping experience, it has become a common practice for online merchants to enable their customers to review or to express opinions on the products that they have purchased. With more and more common users becoming more comfortable with the Web, an increasing number of people are writing reviews. With the growth of online social networking sites, for example, forums, review sites, blogs, and micro blogs, the enthusiasm towards opinion mining has expanded essentially. Today online opinions have transformed into a sort of virtual profit for business organizations looking to market their items, recognize new trends and deal with their position.

Many organizations are currently utilizing opinion mining systems to track customer inputs in online shopping sites and review sites. Opinion mining is additionally helpful for organizations to analyse customer opinions on their products and features. While product attributes are clearly mentioned, discovering the primary cause behind low profit needs much focus on all the more on individual customer views on such characteristics. From these reviews customer can obtain first hand assessment of product information.

Meanwhile the manufacturers can obtain immediate feedback and opportunities to improve the quality of their products. Opinion mining is an amazing method for taking care of numerous business trends identified with deals administration, status management, and advertising. It is a common practice that merchants selling products on the Web ask their customers to review the products and associated services. So opinion mining from online review has become important.

A semi-supervised word alignment model is used to capture opinion relations in sentences, after that, a large number of word pairs, each of which is composed of a noun and its modifier are obtained. Then associations between opinion target candidates and opinion word candidates as the weights on the edges are calculated. The formulation of opinion relation identification is a word alignment process. To employ the word-based alignment model the monolingual word alignment is employed, which has been widely used in many tasks such as collocation extraction.
According to the confidences of all words in data a threshold value is fixed and accordingly the words with higher confidence are assigned as opinion targets while the less ones are opinion words.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

II. RELATED WORK

In [1] L.Zhang and S.H.Lim proposed a method to deal with the problems of the state-of-the-art double propagation method for feature extraction. It first uses part-whole and "no" patterns to increase recall. It then ranks the extracted feature candidates by feature importance, which is determined by two factors: feature relevance and feature frequency.

In [6] Q.Mei, X.Ling, M.Wodra and C.Zhai proposed a probabilistic topic sentiment mixture model (TSM). With this model,the aspects like to learn general sentiment models, extract topic models orthogonal to sentiments, which can represent the neutral content of a subtopic and extract topic life cycles and the associated sentiment dynamics could be effectively studied.

In [8] Z.Liu, H.Wang, H.Wu and S.Li proposed and evaluated a latent document re-ranking method for re-ordering the initial retrieval results. The key to refine the results is finding the latent structure of "topics "or "concepts" in the document set, which leverages the latent Dirichlet allocation technique for the query-dependent ranking problem and results in state-of-art performance.

In [7] A.Mukherjee and B.Liu proposed the problem of modelling review comments, and presented two models TME and ME-TME to model and to extract topics and various comment expressions. The expressions were used to classify comments more accurately, and to find contentious aspects and questioned aspects. The information was used to produce a simple summary of comments for each review.

In [3] Liu et al. focused on opinion target extraction based on the WAM. They used a completely unsupervised WAM to capture opinion relations in sentences. Next, opinion targets were extracted in a standard random walk framework. Liu's experimental results showed that the WAM was effective for extracting opinion targets. Nonetheless, they present no evidence to demonstrate the effectiveness of the WAM on opinion word extraction

Furthermore, a study employed topic modelling to identify implicit topics and sentiment words by Ivan Titov and Ryan McDonald [13]. The aims of these methods usually were not to extract an opinion target list or opinion word lexicon from reviews. Instead, they were to cluster for all words into corresponding aspects in reviews. These methods usually adopted coarser techniques, such as frequency statistics and phrase detection, to detect the proper opinion targets/words. The emphasis is more on how to cluster these words into their corresponding topics or aspects.

III. FRAMEWORK

The framework for proposed system consists of three main processing steps which are Pre-processing, Opinion word and opinion target candidate calculation and Semi-supervised word alignment model. The first step of the system is Pre-processing. Several pre-processing steps are applied on the given reviews to optimize the further operations efficiently. The first step includes splitting of large paragraphed reviews into sorted sentences. Next the tokenization of sentences is carried out, which splits the sentences into sequence of tokens. These result into single token of words. Then is the POS Tagging step. This step consists of labeling the word in a text as corresponding to a respective part of speech. This is based on both its definition and its context i.e. its relationship with adjacent words in a sentence or paragraph.

In opinion word and target candidate calculation process first the confidence of each candidate with respect to the data is calculated using its occurrence of the word in a text. Next the relations between the candidates are identified by assigning a unique number for the each word in a text. The number is repeated when the word reoccurs in the text. Next the word pairs of opinion word candidate and opinion target candidate are obtained by its influence in the neighborhood words and their probabilities.

The Semi-supervised word alignment aims to improve the accuracy of automatic word alignment by incorporating full or partial alignments acquired from humans. Manual correction of such informative links can then be applied to create a labeled dataset used by a semi-supervised word alignment model.

Here, the partial alignment links are regarded as constraints for the trained alignment model. After mining the opinion associations between opinion target candidates and opinion word candidates, a complete Opinion Relation Graph is constructed. Then the confidence of each opinion target/word candidate on this graph is calculated, and the candidates with higher confidence than a threshold are extracted as opinion words as shown in Figure 1.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

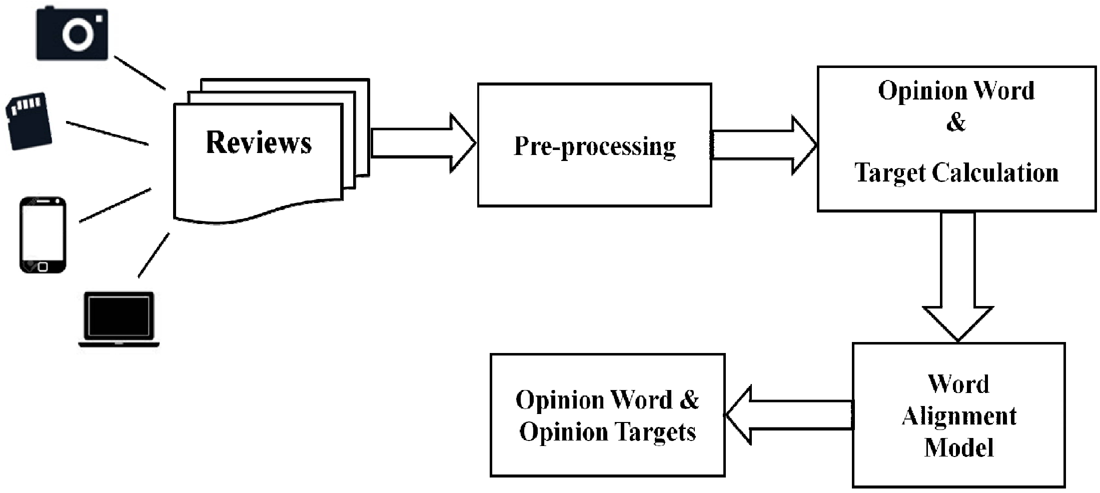

Figure 1: System Framework.

IV. ALGORITHM

Algorithm: Semi-supervised Word Alignment

Unlabeled Data Set : $D = (s_k, t_k)$
Alignment Set : $M_0 = \{a^k_{ij}, \forall_{si} \in S_k, t_i \in T_k\}$
Train Semi-supervised Word Alignment using
$(D, M_0) \rightarrow \alpha_0$
for each word in T **do**
L_t = LinkSelection(D, M_t, α_t, N)
Request Human Alignment for L_t
$M_{t+1} = M_t + L_t$
Re-train Semi-Supervised Word Alignment
on $(D, M_{t+1}) \rightarrow \alpha_{t+1}$
end for

V. MATHEMATICAL MODEL

The alignment probabilities between an opinion target (w_t) and a potential opinion word (w_o) are estimated using

$$P(wt|wo) = \frac{Count(wt,wo)}{Count(wo)}$$

Where, P(wt|wo) means the alignment probability between opinion target (w_t) and a potential opinion word (w_o). Similarly, the alignment probability P(wo|wt) is obtained by changing the alignment direction in the alignment process.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

VI. RESULTS

We use the customer review collection is used as the testing data. The dataset is the Customer Review Datasets (CRD), which includes English reviews of five products. The collection contains five review data sets: two on two digital cameras, one on a DVD player, one on an mp3 player, and one on a cell phone. It consists of dataset, domain, language and number of sentences, number of opinion words and number of opinion target. We select precision (P), recall (R) and F-measure (F) as the evaluation metrics.

The proposed Semi-Supervised Word Alignment algorithm provides more precise results compared to partially supervised word alignment model, since the human alignments are used to examine the words more accurately. This makes the system more efficient to extract the opinion targets and opinion words.

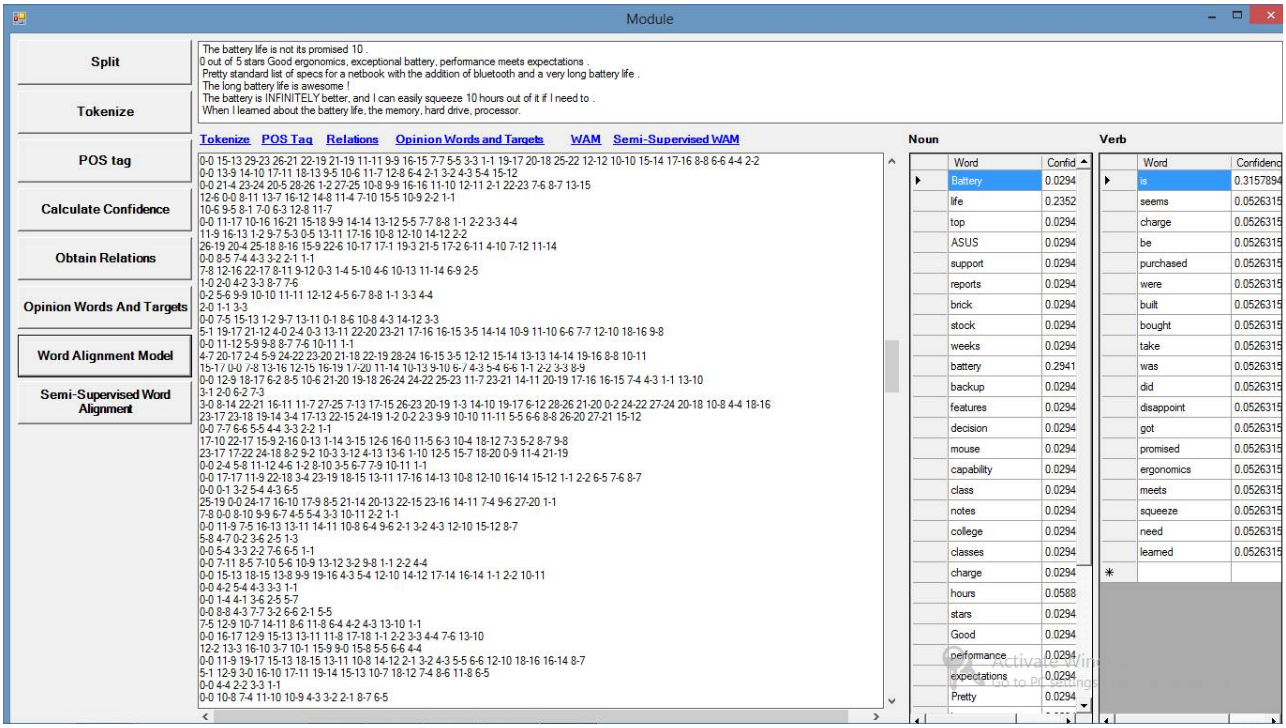

Figure 2: Proposed System.

VII. CONCLUSION

Opinion mining has become an increasing activity in web. Having right product information about a product is important to present before the right customer. With rapid development of web, a huge number of product reviews are springing upon the web. From these reviews customer can obtain first hand assessment of product information. Meanwhile the manufacturers can obtain immediate feedback and opportunities to improve the quality of their products.

Here a unique method for extracting the opinion targets and opinion words by using a Semi-supervised word alignment model is proposed. The vital contribution is focused on detecting opinion relations between opinion targets and opinion words. This method captures the opinion relations more precisely compared to partially-supervised word alignment model and therefore the opinion target and opinion word are effectively extracted. Next, an Opinion Relation Graph to model all candidates and the detected opinion relations among them is constructed with a graph co-ranking algorithm to estimate the confidence of each candidate. The items with higher ranks are extracted out.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

REFERENCES

[1] Kang Liu, Liheng Xu, and Jun Zhao, "Co-extracting opinion targets and opinion words from online reviews based on the word alignment model" in IEEE Transactions on Knowledge and Data engineering, vol. 27, no. 3, March 2015,pp. 636-650
[2] 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.
[3] L. Zhang, B. Liu, S. H. Lim, and E. O'Brien-Strain, "Extracting and ranking product features in documents," in Proc. 23th Int. Conf. Comput.Linguistics, Beijing, China, 2010, pp. 1462–1470.
[4] 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. 1346–1356.
[5] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in Proc. Conf. Web Search Web Data Mining, 2008, pp. 231–240.
[6] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai, "Topic sentiment mixture: modeling facets and opinions in weblogs," in Proc. 16th Int. Conf. World Wide Web, 2007, pp. 171–180.
[7] A. Mukherjee and B. Liu, "Modeling review comments," in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Jeju, Korea, Jul. 2012, pp. 320–329.
[8] Z. Liu, H. Wang, H. Wu, and S. Li, "Collocation extraction using monolingual word alignment method," in Proc. Conf. Empirical Methods Natural Lang. Process., Singapore, 2009, pp. 487–495.
[9] 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.
[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. 653–661
[11] 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.
[12] 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.
[13] I. Titov and R. McDonald, "A joint model of text and aspect ratings for sentiment summarization," in Proc. 46th Annu. Meeting Assoc. Comput. Linguistics, Columbus, OH, USA, 2008, pp. 308–316.
[14] 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.
[15] 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.

BIOGRAPHY

Soha N. Chaudhari received the B.E. degree in Information Technology from MET's Institute of Engineering, Nashik. Currently pursuing M.E. degree in Computer Engineering at SVIT, Chincholi, Nashik.

Manisha M. Naoghare received the B.E. degree in Computer Engineering and M.E. in Computer Science Engineering from P.R.M.I.T&R, Badnera, Amravati. Currently she is working as a Asst.Professor at SVIT, Chincholi, Nashik in Computer Engineering Department.