ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

Opinion Targets and Word Extraction Using Word Alignment Model

[1] Apoorva A. Lonkar, [2] Savita Sangam,
[1] UG Student, Cummins College of Engineering for Women Karve Nagar, Pune
[2] Head of Department (IT), Shivajirao S. Jondhale College of Engineering, Dombivli (E), Thane

Abstract: **- Opinion mining is important in learning human behaviors and different personality traits by extracting information from all possible instances such as hidden emotions, unidentified representative, big texts, and even the use of urban language. The fastest growing social media forum such as the Facebook allows all its users to contribute information without borders; sharing opinions about current world issues, and even their attitude and views towards life. Thus, opinion mining is one of most well-known and important fields of study nowadays. This report analyses different approaches used for opinion mining. It explains the stages of analysis and the tools used for it. This study demonstrates partially-supervised Word Alignment Model for sentiment classification of text reviews. Using this model mining opinion relations between words under constraints can be developed. The partially-supervised word alignment model identifies opinion relations as an alignment process. The proposed model obtains better accuracy because of the use of partial supervision as compared to the supervised or unsupervised alignment model.**

Keywords**: - Opinion mining, partially-Supervised Word Alignment Model, Sentiment Classification, Word alignment Model.**

I. INTRODUCTION

What is the opinion of the other people having always been an important piece of information for most of us during the decision-making process. More and more people are making their opinions available to strangers via the Internet. Thus large amount of data is getting produced every second in the form of reviews, opinions through social media. This is semi structured/ unstructured in nature and need to be processed, analyzed to get the useful information. The process that does these tasks is termed as opinion mining in the literature. It is part of text mining, a branch of data mining. Figure 1.1 shows the Data Mining hierarchy.

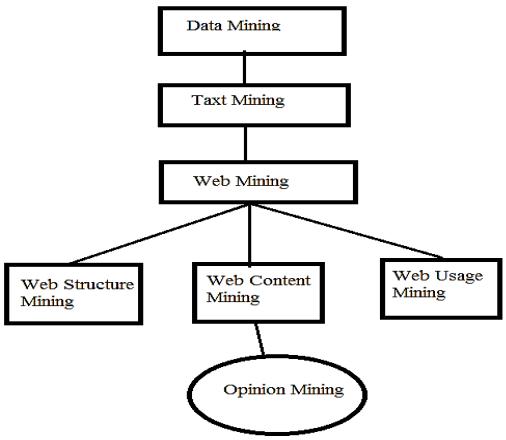

Figure1.1 Data Mining Hierarchy

Opinion mining is the study of analyzing people's viewpoint towards products, companies or about other people from their written review (language). The terms opinion mining and sentiment analysis can be used interchangeably. A huge volume of opinionated data recorded in digital form is now available for analysis. Opinion mining consists of attempting to ascertain the sentiment that is positive or negative polarity of a text passage by analyzing the features of that passage. A user can be trained to classify new instances by analyzing the text features associated with different opinions along with their sentiment. Constructing features to describe the text of the review involves text mining and Natural Language Processing (NLP). Sentiment Analysis is an application of natural language processing, computational linguistics, and text analytics to identify subjective information from source data. It clubs the sentiments into categories like "positive" or "negative". Thus, it determines the general attitude of the speaker or a writer with respect to the topic in context. Sentiment analysis/opinion mining systems are being applied in almost every business and social domain because opinions are central to almost all human and are key influencers of our behaviors. The greatest challenge of sentiment analysis is to design application-specific algorithms and techniques that can analyses the human language linguistics accurately.
Following are the major applications of sentiment analysis in real world scenarios.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

1. Product and Service reviews - The most commonly used application of sentiment analysis is in reviews of consumer products and services.
2. Reputation Monitoring - Twitter and Facebook are a focal point of many sentiment analysis applications.
3. Result prediction - One can predict the probable outcome of an event, by analyzing sentiments from relevant sources. Sentiment analysis can provide substantial value to candidates running for various positions.
4. Decision making – It can be used as an important factor assisting the decision-making systems such as financial markets investment.

In this work, an attempt has been made to understand the tasks of Opinion Mining by going through the work published by various researchers. The detailed literature survey is given the next section

II. LITERATURE SURVEY

Minqing Hu and Bing Liu in [1] propose a unique technique to perform sentiment classification. Their task is mainly performed in three steps: 1) mining product features 2) identifying opinion sentences and deciding whether each opinion sentence is positive or negative 3) summarizing the results. They also proposed a set of techniques for mining and summarizing product reviews based on data mining and natural language processing methods.

The authors in [2] also suggests that they have designed several syntax-based patterns to capture opinion relation in sentences, and used a bootstrapping algorithm (called Double Propagation) to extract opinion targets and opinion words. This paper is based on two important tasks in opinion mining that are Opinion lexicon expansion and target extraction. They propose a propagation approach to extract opinion words and targets iteratively given only a seed opinion lexicon of small size. The extraction is done using identified relations between opinion words and targets.

The paper [3] is an extension of double propagation algorithm. Besides the syntactic pattern used in double propagation, zhang designed some heuristic pattern to indicate opinion target candidates. Feature extraction for entities is an important task for opinion mining. The paper proposed a method to deal with the problems of the state-of-the-art double propagation method for feature extraction. In that method, it first uses part-whole pattern and “no” patterns to increase recall. It then ranks the extracted feature candidates according to feature importance, which is determined by two factors: feature relevance and feature frequency. The Web page ranking algorithm HITS was applied to compute feature relevance.

In the review paper [4] authors studied existing extracting opinion targets and opinion words system. Word alignment model includes heavy tasks, hence to balance the load and prepare the execution process partial supervised technique is used and syntactic patterns are used for it. This opinion mining concept is also integrated in web based application that suggests products based on the analysis of product reviews. Previously existed system faced problem such as, they use nearest-neighbor rules for nearest adjective/verb to a noun/noun phrase in a minimum window as its modifier, resulting in less accurate results. They have used opinion relation graph to detect several opinion relations of candidates. In this paper [5] authors replace the previous method with Word Alignment Model and Topical Word Trigger Model which are tested for results. This paper proposes an innovative approach to conjointly extract them with graph co-ranking. Their co-ranking is customized and every candidate’s confidence is merely determined by its most well-liked collocations. It helps to boost the extraction exactitude. These models capture opinion relations more exactly and therefore are more effective for opinion target and opinion word extraction.

III. OPINION MINING

Opinion mining is a type of text mining for tracking the mood of the public. Some of the approaches for opining mining are discussed here.

3.1. Different approaches for opinion mining

1. Frequency-based approach:

If user want may want to know about the quality of the food in a hotel, but reviewers may focus on other aspects of the hotel, such as the decor or the location. Hence, reader is forced to wade through many reviews looking for information about features of interest.

The main subtasks for this approach:

a) Identify features associated with the product
b) Identify opinions regarding product features.
c) Determine the polarity of opinions as positive and negative.
d) Rank opinions based on their strength.

OPINE [7], an unsupervised information extraction system can be used to solve all these tasks. OPINE uses the frequency based point of view for opinion mining. It mines reviews to build a model of important product feature. Given a product and a corresponding set of reviews, OPINE solves the opinion mining tasks indicated above and outputs a set of product features, each accompanied by a list of associated opinions which are ranked based on strength. This output information can then be used to produce various types of opinion summaries. OPINE uses association rule mining to extract frequent review noun phrases as features.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

2. Relation-based approach:

1] Opinion Observer: A prototype system called Opinion Observer uses the relation-based approach for opinion mining. In this, at one glance of its visualization, the user can identify the strengths and weaknesses of each product in the minds of consumers in terms of various product features. Both potential customers and product manufacturers can benefit from this comparison.

2] Multi-facet rating: Software tools to organize product reviews and to make them easily accessible to prospective customers are going to be more and more popular. Some of the issues that the designers of these software tools need to address are pulling together reviews from various resources, filtering our fake reviews given by authors with vested interests and ranking products automatically products in terms of the satisfaction of consumers that have purchased the product before.

3. Supervised learning:

The Opinion Miner system used the organized approach for opinion mining. Some merchants who sell their products on the Web ask their customers to share their opinions and hands-on experiences on products they have purchased.

The Opinion Miner system is designed with a aim to mine customer reviews of a product and extract high detailed product entities on which reviewers express their opinions. This system first locates opinion expressions and then opinion intentions for each accepted product entity are classified as positive or negative.

4. Topic modelling:

Topic-Sentiment Mixture (TSM) is an opinion mining model which uses the topic modelling approach. It is a probabilistic model. This model addresses the Topic-Sentiment Analysis (TSA) problem and extracts the many subtopics and sentiments in a collection of blog articles. A blog article is examined to be "generated" by sampling words from a mixture model. The mixture model contains a background language model, topic language models, and two positive and negative sentiment language models. By using this model, we can extract the topic/subtopics from blog articles, reveal the correlation of these topics and different sentiments, and further model the dynamics of each topic and its associated sentiments.

3.2. Tools used for Opinion Mining

The tools used to trace the opinion or polarity from the user generated contents are:

- Review Seer tool [8] – This tool is used to automates the work done by aggregation sites. The Naive Bayes classifier approach is used to collect positive and negative opinions for allocating a score to the extracted feature terms.
- Web Fountain [9] - It uses the beginning definite Base Noun Phrase (bBNP) heuristic approach for extracting the product features.
- Opinion observer [8]-This is an opinion mining system for examining and comparing opinions on the Internet using user generated contents. This system shows the results in a graph format showing opinion of the product feature by feature. Along with these automated tools, there are various online tools like Twitrratr, Twendz, Social mention, and Sentimetrics are available to trace the opinions in the web. This works demonstrates Word Alignment Model for sentiment classification of reviews/text. The method is described in detail in the next section.

IV. PROPOSED METHOD

4.1 Word Alignment Model:

We employ the word-based alignment model [6] to perform monolingual word alignment, which has been popularly used in many tasks such as collocation extraction and tag suggestion. In general, every sentence is duplicated to generate a parallel corpus. A bilingual word alignment algorithm is used on the monolingual scenario to align a noun with its modifiers (potential opinion words) in sentences. An opinion relation graph is shown in figure 4.1.

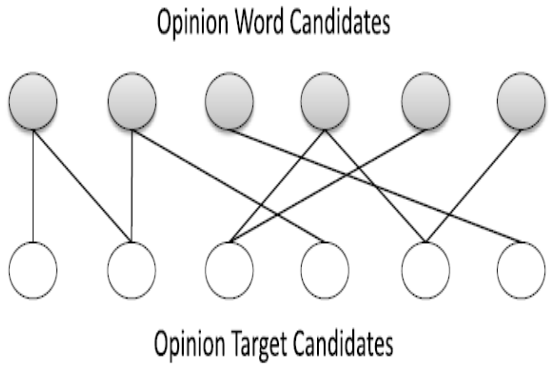

Figure 4.1 Opinion relation graph

In this process, we are extracting opinion targets/words as a co-ranking method. We think that all nouns/noun phrases are opinion target candidates, and all adjectives/verbs are regarded as potential opinion words. Each candidate is assigned with a confidence, and candidates with confidence greater than threshold are extracted as the opinion targets or opinion words. To assign a confidence to each candidate, our basic motivation is as follows.

If a word is likely to be an opinion word, the nouns/noun phrases with which that word has a modified relation will have higher confidence as opinion targets.

If a noun/noun phrase is an opinion target, the word that modifies it will be highly likely to be an opinion word.

We can infer that the confidence of a candidate (opinion target or opinion word) is collectively determined by its neighbors as per the opinion associations among them. An example is shown in figure 4.2.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

Example:

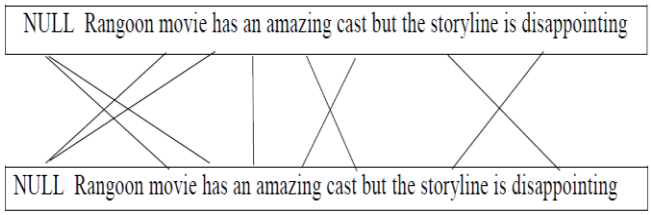

Figure 4.2. Mining opinion relations between words using the Word Alignment Model

We obtain the above alignment results as shown, where "NULL" refers to the null word. In this example, we can see that unrelated characters, such as "is", "has", "an", align themselves. There are no opinion words to modify such as "movie" and "has" therefore, these two words may align with "NULL". There are several word alignment models for usage, namely IBM-1, IBM-2 and IBM-3. We have selected IBM-3 model in this experiment. For training the IBM-3 model, the simpler models are sequentially trained to become the initial alignments for the upcoming/ subsequent model.

Constraints of word alignment model:
Nouns (adjectives/verbs) must be aligned with adjectives/verbs (noun phrases) or a null word. Aligning to a null word means that this word either has no modifier or modifies nothing. Other unrelated words, like prepositions, conjunctions and adverbs, can only align themselves.

4.2 Partially-Supervised Word Alignment Model:
The word alignment model has proven to be efficient for opinion target extraction process. But, for the same, there is still no evident proof to demonstrate the Word Alignment Model's correctness. We also notice that standard word alignment models are usually trained in a completely unsupervised manner, which results in alignment quality that may prove to be disappointing. We can improve alignment quality by using supervised approaches. However, it is both time consuming and impractical to manually label full alignments in sentences. Thus, we employ a partially-supervised word alignment model (PSWAM) for the betterment of alignment performance. We perform a partial supervision on the statistic model and employ a partially-supervised alignment model to incorporate partial alignment links into the alignment process. Here, the partial alignment links are treated as constraints for the trained alignment model.
Formally, given the partial alignment links Â= {(i,ai)|i ϵ [1, n],ai ϵ [1, n]}, the optimal alignment A* in eq. (1) is rewritten as follows:

$$A^* = argmax\ P(A|S, \hat{A}).$$

An example is shown here in Figure 4.3.

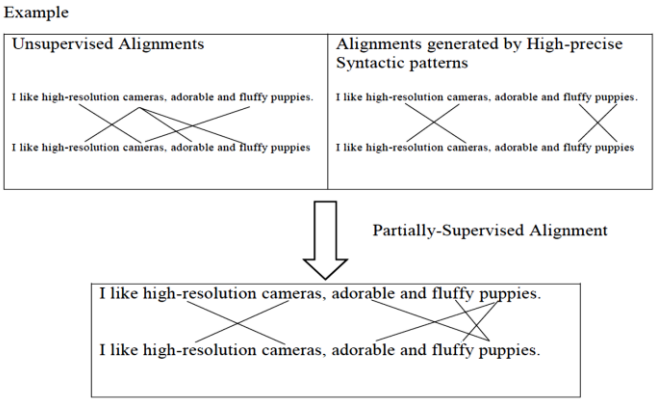

Figure 4.3. Mining opinion relations between words using partially supervised alignment model

V. CONCLUSION

An attempt has been made to understand the concepts of Opinion mining especially sentiment classification which is a sub task in opinion mining. A detailed study of various approaches for opinion mining is undertaken and found that word alignment model is one such approach for sentiment classification. We also notice that standard word alignment models are usually trained in a completely unsupervised manner, which results in alignment quality that may prove to be unsatisfactory. We can improve alignment quality by using supervised approaches. However, it is both time consuming and impractical to manually label full alignments in sentences. Thus, we employ a partially-supervised word alignment model (PSWAM) for sentiment classification. This model yields better classification results when compared to the supervised or unsupervised word alignment models.

REFERENCES

1. 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.

2. 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.

3. 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.

4. R.S. Gaikwad, Dr. B.L.Gunjal "A Review on Existing Opinion Targets and Opinion Words Extracting Methods from Online Reviews" Vol-2 Issue-1 2016 IJARIIE-ISSN (O)-2395-4396.

5. Utkarsha Vibhute, Prof. Soumitra Das "Pulling out opinion targets and opinion words from reviews based on the word alignment model and using topical word trigger model"Multidisciplinary Journal of Research in Engineering and Technology, Volume 2, Issue 4, Pg.834-840.

6. P. F. Brown, V. J. D. Pietra, S. A. D. Pietra, and R. L. Mercer, "The mathematics of statistical machine translation: Parameter estimation," Comput. Linguist. vol. 19, no. 2, pp. 263–311, Jun. 1993.

7. S. Moghaddam and M. Ester, "Aspect-based opinion mining from online reviews," in Proc. Tutorial 35th Int. ACM SIGIR Conf., New York, NY, USA, 2012.

8. G.Vinodhini and RM. Chandrasekaran 2012. Sentiment analysis and Opinion Mining: A survey International Journal of advanced Research in Computer Science and Software Engineering vol. 2 Issue 6.

9. Yi and Niblack 2005. Sentiment Mining in Web Fountain", Proceedings of 21st international Conference on Data Engineering, pp. 1073-1083, Washington DC.