

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 6.017

IJCSMC, Vol. 8, Issue. 1, January 2019, pg.38 – 49

SURVEY OF CO-EXTRACTING OPINION TARGETS AND OPINION WORDS FROM ONLINE REVIEWS BASED

Malathi.M¹, Dr. Antony Selvadoss Thanamani²

Ph.d Research Scholar, Dr. Mahalingam Centre for Research and Development, NGM College, Pollachi, India¹

Associate Professor and Head, Dr. Mahalingam Centre for Research and Development, NGM College, Pollachi, India²

¹ malathi.mca09@gmail.com, ² selvdoss@gmail.com

***ABSTRACT:** Due to the sheer volume of opinion rich web resources such as discussion forum, review sites, blogs and news corpora available in digital form; much of the current research is focusing on the area of sentiment analysis. People are intended to develop a system that can identify and classify opinion or sentiment as represented in an electronic text. An accurate method for predicting sentiments could enable us, to extract opinions from the internet and predict online customer's preferences, which could prove valuable for economic or marketing research. Till now, there are few different problems predominating in this research community, namely, sentiment classification, feature based classification and handling negations. This paper presents a survey covering the techniques and methods in sentiment analysis and challenges appear in the field.*

I. INTRODUCTION

Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product or topic. Sentiment analysis, which is also called opinion mining, involves in building a system to collect and examine opinions about the product made in blog posts, comments, reviews or tweets. Sentiment analysis can be useful in several ways. For example, in marketing it helps in judging the success of an ad campaign or new product launch, determine which versions of a product or service are popular and even identify which

demographics like or dislike particular features. There are several challenges in Sentiment analysis. The first is a opinion word that is considered to be positive in one situation may be considered negative in another situation. A second challenge is that people don't always express opinions in a same way. Most traditional text processing relies on the fact that small differences between two pieces of text don't change the meaning very much. In Sentiment analysis, however, "the picture was great" is very different from "the picture was not great". People can be contradictory in their statements. Most reviews will have both positive and negative comments, which is somewhat manageable by analyzing sentences one at a time. However, in the more informal medium like twitter or blogs, the more likely people are to combine different opinions in the same sentence which is easy for a human to understand, but more difficult for a computer to parse. Sometimes even other people have difficulty understanding what someone thought based on a short piece of text because it lacks context. For example, "That movie was as good as its last movie" is entirely dependent on what the person expressing the opinion thought of the previous model. The user's hunger is on for and dependence upon online advice and recommendations the data reveals is merely one reason behind the emerge of interest in new systems that deal directly with opinions as a first-class object. Sentiment analysis concentrates on attitudes, whereas traditional text mining focuses on the analysis of facts. There are few main fields of research predominate in Sentiment analysis: sentiment classification, feature based Sentiment classification and opinion summarization. Sentiment classification deals with classifying entire documents according to the opinions towards certain objects. Feature-based Sentiment classification on the other hand considers the opinions on features of certain objects. Opinion summarization task is different from traditional text summarization because only the features of the product are mined on which the customers have expressed their opinions. Opinion summarization does not summarize the reviews by selecting a subset or rewrite some of the original sentences from the reviews to capture the main points as in the classic text summarization

II. RELATED WORKS

2.1 An approach of Opinion mining for online marketing Using Sentiment Thesaurus and Concept Search Engine

Emotion analysis, also known as opinion mining, provides techniques for analyzing and identifying the sentiment expressed in the blog post. The sentence in the blog post can be categorized as objective and subjective sentence. The Subjective Sentence plays a major role in expressing the opinion and the objective sentence deals with facts. So we have to consider only the subjective sentence for predicting the emotion. Extraction of the relevant feature plays a major role for exact opinion mining. Previously the feature was extracted based on the number of occurrence of the term. The Drawback in that is the number of features may be more which will complicate the application.

Customer Reviews for Product and allied services has become a de-facto for Vendors selling products online. Customer reviews for a product is becoming more important day by day as e-commerce industry grows imminently. For a renowned product, the number of customer reviews can run to many counts. This makes the buying decision for the customer more complex. The crux of this research work is to do a summarization of the entire customer reviews of product. This summarization task calls out the specific feature details like opinions of the product unlike the conventional text summarization including positive and negative. No original sentences of reviews are summarized by selecting or rewriting to identify the important concepts as in the classic text summarization.

It has focused on unsupervised classification, but average semantic orientation could be augmented by other typical concepts, in a supervised classification system. The other features could be based on the presence or absence of specific words to yield higher precision. This simplifies the analysis, before combining it with other features. We then utilize the made thesaurus to extend characteristic vectors amid train also test times for a paired classifier. An important sub set of the gimmicks is chosen utilizing LI regularization. The proposed system altogether outflanks a few baselines and reports come about that are practically identical with a while ago proposed cross-area notion grouping routines on benchmark information set. Also, our examinations against the Senti word net demonstrate that the made supposition delicate thesaurus Precisely gatherings words that express comparative suppositions.

2.2 Cross-Domain Sentiment Analysis of Product Reviews by Combining Lexicon-based and Learn-based Techniques

Text sentiment analysis is a hot research topic in recent years in terms of the Natural Language Processing. There are mainly two kinds of sentiment analysis methods: using sentiment lexicon to extract affective elements as the classification basis; using machine learning methods for sentiment classification.

Many techniques have been proposed to tackle the text sentiment classification problem which could be classified into two categories: lexicon-based and learn-based techniques. They either need labeled sentiment lexicons or require a large amount of labeled corpus as a training set. Besides, lexicon based algorithms rely on sentiment lexicon too much, and some words carrying emotional information are not included in the sentiment lexicon. What's more, sentiment is expressed differently in different domains. For example, in the domain of electronics, the words "heat" and "runs fast" are used to express positive sentiment, however, in the book domain, the words "heat" and "runs fast" express none sentiment. Presented a lexicon-based approach for extracting sentiment from text.

Product reviews can direct consumers' purchasing behaviors and sellers' marketing strategies. Therefore, in this paper, we propose a novel sentiment analysis method which Combines Lexicon-based and Learn-based techniques (CLL) to analyze the cross-domain sentiment of Chinese product reviews. We first build three domain lexicons based on the basic lexicon and corpus from three domains containing books, hotels and electronics. Furthermore, we use four categories of features (including 16 features in total) to build six classifiers. We conduct a series of experiments to evaluate our proposed CLL by using different lexicons and different classifiers.

A novel framework CLL method. Specific lexicons are built for each domain, and the results of lexicon-based method are regarded as input features of the learn-based method. Compared with existing algorithms, our method has four main contributions: Achieve three domain-adaptive sentiment lexicons for domains of books, hotels and electronics. Provide a sentiment feature set including semantic information and structural information. Rank features by their information entropy for further understanding the importance of these features in different domains. Propose CLL method which performs better than state-of-the-art methods in domains of books and hotels.

2.3 OpinMiner: Extracting Feature-Opinion Pairs with Dependency Grammar from Chinese Product Reviews

Opinion mining (or product review mining) has been extensively studied in recently years. A majority of these researches have focus on identifying the polarity expressed in various opinion units such as word, phrase, sentence or review documents. While not so much work has been done on feature level opinion mining, especially for Chinese review. With the flourish of the Web, online review is become a more and more useful and important information resource for people. As a result, automatic review mining has become a hot research topic recently. Traditional review mining based on feature extracts product feature and opinion word independently and seldom considers their association information. In this paper, we only focus on Chinese product review. We propose a method based on Chinese dependency grammar to extract feature-opinion word pairs. Specifically, we use Chinese dependency grammar to set several rules, and then we make use of these rules to extract candidate feature-opinion word pairs.

The main idea of this grammar is that roles played by words of different grammar roles in a sentence are not equal to each other, saying that, some words depend on others while some words govern others. We take use of the dependency parsing module in the HIT LTP shared library (<http://ir.hit.edu.cn>) to do the parsing. We make use of Chinese dependency grammar to describe to relations syntactically. In the dependency grammar, a syntactic relation between two words A and B can be described as A (or B) depends on B (or A). We define two categories to summarize all possible dependencies between two words in sentences.

Then, we use these rules to extract candidate feature-opinion pairs directly. Finally, we filter out mismatched feature-opinion pairs by feature ranking and Named Entity Recognition (NER) system. In addition, we consider the situation that noun phrase is broken up into two nouns because of false Chinese segmentation, which make us extract mismatched feature opinion word pairs. Colloquial nature of Chinese product reviews and false Chinese Segmentation make the effect of extraction a little terrible. We identify feature by extracting feature-opinion pairs, This can improve the precision of feature detection. Feature-opinion pairs impose restrictions on feature so that feature and opinion word co-occur in the same sentence by certain relation.

We will explore how to further improve the precision and the recall of feature-opinion pairs. At the same time, we will consider the situation that a feature matches several opinion words, several features match a opinion word and no feature occurs in a review sentence.

2.4 Mining Consumer's Opinion Target Based On Translation Model And Word Representation

To capture the user's opinion target, many researchers focus on opinion word as an indicator. With the syntax-based method used to get the associations between opinion target and opinion word, classic opinion target extraction method often deploy one of the syntax parser like Stanford NLP parser, results of such parsers act as important features, this is an effective way in specific corpus. However, the customer reviews usually organized less formally, especially the reviews in Chinese often have word segment problems, which will significant effect the precision and recall of the extraction result.

E-commercial plays an important role in people's daily life. When on the Internet, people often buy commodities from Taobao, Tmall and make comments on them, the comments of the goods may have closely connections with commercial value, which often reflect what's the consumers really care when they choose one piece of good among thousands of other similar ones. How to mining these aspects which the consumers really concern is a problem left unsolved. As a potential effective solution to construct structured information for people's preference, Information Extraction (IE) has attracted more and more scholar's attention. A meaningful research area is Opinion Target Extraction (OTE). This paper proposed a system using translation model as well as word representation method to obtain user's interests on dataset in Chinese. To release the word segmentation error, a finely generated system with new Chinese word detection module is proposed. The experiments on two corpus subjected on digital product verify the effective of our method.

For Chinese reviews, we proposed a module which can effectively find out the new word, and the important Chinese New Word detection will significantly improve the result. Translation model show remarkable ability in IE, however, the word alignment is based on word frequency and other information, so we introduce the semantic method into the corpus and get better performance. The experiments on two corpus subjected on digital product verify the effective of

our method. We add the semantic information into the system to release the noise of word frequency's influence from the word aliment model. Then we use HITS algorithm to re-rank the Node as the top opinion target.

2.5 Recommending Products to Customers using Opinion Mining of Online Product Reviews and Features

E-Commerce and E-Retailing markets with the market expected to grow to around USD 9 billion by 2016. With such a rapid growth in this industry, companies are using sophisticated algorithms to understand the buying patterns of their buyers in order to enrich the customer experience. There is cut throat competition among Ecommerce sites in the way they present their products, the promotions and discounts they offer and the shopping experience they provide to customers. These offerings are based on extensive market research and analytics conducted by experts within and outside these organizations. One of the key parameters that companies use to strategize is customer reviews and rating on the e-commerce sites. These reviews are not only used by the companies but also play a major role in consumers deciding whether to buy a product or not. Hence analyzing customer reviews help both shoppers as well as E-Commerce companies.

Online customer reviews are helping consumers to decide which products to buy and also companies to understand the buying behavior of consumers. In this paper we have created a prototype web based system for recommending and comparing products sold online. We have used natural language processing to automatically read reviews and used Naive Bayes classification to determine the polarity of reviews. We have also extracted the reviews of product features and the polarity of those features.

Opinion mining is rapidly evolving research area with newer and newer technologies and algorithms enabling the automatic processing of data. Opinion mining is the study of people's sentiments and opinions about objects and the various aspects of the objects. Opinion mining has Several applications, some of which include: Understanding the attitudes and sentiments of people on social networks like facebook and twitter. Understanding the buying intention of customers on e-commerce sites. Understanding the satisfaction of the customer with specific products through online reviews on e-commerce sites. Recommending products and places to customers based on collective opinions. We have also done a feature based classification of

reviews. The objective is to benefit the customers and assist them in choosing the right product. Natural language processing (NLP) techniques to tell if the opinion expressed by the reviewer is positive or negative. NLP techniques also help in extracting the portions of reviews where reviewers talk about specific product features.

2.6 A Logic Programming Approach to Aspect Extraction in Opinion Mining

Aspect extraction aims to extract fine-grained opinion targets from opinion texts. Recent work has shown that the syntactical approach performs well. In this paper, we show that Logic Programming, particularly Answer Set Programming (ASP), can be used to elegantly and efficiently implement the key components of syntax based aspect extraction. Specifically, the well known double propagation (DP) method is implemented using 8 ASP rules that naturally model all key ideas in the DP method. Our experiment on a widely used data set also shows that the ASP implementation is much faster than a Java-based implementation. Syntactical approach has its limitation too. We identify a set of general words from WordNet that have little chance to be an aspect and prune them when extracting aspects. The concept of general words and their pruning are concisely captured by 10 new ASP rules, and a natural extension of the 8 rules for the original DP method. Experimental results show a major improvement in precision with almost no drop in recall compared with those reported in the existing work on a typical benchmark data set. Logic Programming provides a convenient and effective tool to encode and thus test knowledge needed to improve the aspect extraction methods.

We propose to employ logic programming for effective experimentation and building of the key components of aspect extraction systems. Note that basic factual information on the words in the opinion texts and some routine processing are still implemented using traditional language. In this approach, the syntactic relations and the knowledge on how those relations are related to aspect and opinion words are represented naturally as logical rules. The aspects will then be automatically generated by the logic programming inference engines from the logic rules.

Compute the answer set of the logic program resulted from the first three steps using existing ASP solvers. It is also 10 times faster than a Java implementation on a collection of widely used data sets. We also proposed to augment the DP method with new knowledge for pruning aspects from the candidate ones. Experiments on 5 benchmark review data sets show

that the augmented DP method outperforms the state-of-the-art DP method by a large margin. In future work, we plan to investigate more syntactic extraction rules, to refine knowledge for identifying incorrect aspects, and to conduct experiments on more data sets. We also plan to explore if other LP systems provide better performance (in terms of both modeling effectiveness and efficiency) than ASP.

2.7 e - Document Retrieval using Rough-Set Theory

With the growing use of networked solutions and services in the Information Society, many citizen-oriented services are planned not only in utility or e-governance sectors, but also in other information-centric human applications. Conventional DBMS may have either two-tier or three-tier client/server architecture. These are already being used on the Internet. Web Applications provide Network Interfaces to access information stored in one or more databases. The generic method of organizing and formatting these Network Interfaces is through the hyperlink documents, which are coded in languages like HTML or XML. Further, the proliferation of Data Mining and Data Warehousing practices has necessitated the need for efficient e-Document retrieval, creating a requirement for robust Information Retrieval (IR) systems. In this paper we show that rough-set theory characterizes new and interesting viewpoint for electronic document (e-Documents) retrievals. We explore the characteristics of an e-Document and applicability of rough-set theory in detail.

The starting point of the text analysis process may be the complete document, an abstract, the title only, or maybe a list of words. The frequency of word occurrence in an article furnishes a useful measurement of word significance.' He considered 'f' to be the frequency of occurrence of various word types in a given position of text and 'r' their rank order and plotted a graph relating f and r similar to hyperbolic curve. The text analysis process requires generation of document representatives in a way that involves minimum human intervention. This generation process is termed as *conflation*. Such a system will usually consist of three parts: (1) removal of high frequency words, (2) suffix stripping, (3) detecting equivalent stems. Index Term Weighting can also be thought of as a process of normalization, in a case where we may wish to ensure that a match in one term among ten bears more weight than one amongst twenty.

A retrieval process uses Meta language tags or a derived text structure for indexing in the form of an inverted file format. The retrieval process is depicted IR system matches the query

descriptors with the descriptors of the documents in the input database, following which it ranks the retrieved set of documents in order of relevance. The system then reports back to the user. Typically this needs to be viewed from the rough set approach to assess the likelihood of relevance with the concept of the information sought by the user. It is expected that such an approach is useful for academic environment for introducing courseware.

III. CONCLUSION

Sentiment detection has a wide variety of applications in information systems, including classifying reviews, summarizing review and other real time applications. There are likely to be many other applications that are not discussed. It is found that sentiment classifiers are severely dependent on domains or topics. From the above work it is evident that neither classification model consistently outperforms the other, different types of features have distinct distributions. It is also found that different types of features and classification algorithms are combined in an efficient way in order to overcome their individual drawbacks and benefit from each other's merits, and finally enhance the sentiment classification performance.

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] 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.
- [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] 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.
- [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. 755–760.
- [6] 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. 339–346.
- [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 cross inducing," in Proc. 3rd Int. Joint Conf. Natural Lang. Process., Hyderabad, India, 2008, pp. 289–295.
- [9] B. Liu, *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*, series Data-Centric Systems and Applications. New York, NY, USA: Springer, 2007.
- [10] 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. 1199–1204.
- [11] 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. 81–88.
- [12] 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.
- [13] 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.
- [14] Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," in Proc. Conf. Empirical Methods Natural Lang. Process., Singapore, 2009, pp. 1533–1541.
- [15] 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.
- [16] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1-2, pp. 1–135, Jan. 2008.
- [17] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, pp. 1–167, 2012.
- [18] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2004, pp. 168–177.
- [19] L. Ku, Y. Liang, and H. Chen, "Opinion extraction, summarization and tracking in news and blog corpora," in *Proceedings of AAAI-CAAW-06, the Spring Symposia on Computational Approaches to Analyzing Weblogs*, 2006.
- [20] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," *Comput. Linguist.*, vol. 37, no. 1, pp. 9–27, Mar. 2011.
- [21] Hu, Liu and Junsheng Cheng, "Opinion observer: analyzing and comparing opinions on theWeb", *Proceedings of 14th international Conference on WorldWideWeb*, pp. 342-351, Chiba, Japan, 2005.
- [22] Jia, C. Yu, and W. Meng, "The Effect of Negation on Sentiment Analysis and Retrieval Effectiveness", In *Proceedings of CIKM*, 2009.

BIOGRAPHY



M. Malathi received her B.C.A and Master of Computer Applications from NGM College, Pollachi, Coimbatore, India. Presently she is working as an Assistant Professor in the Department of Computer Science in NGM College (Autonomous), Pollachi. Her area of interest includes data Mining, opinion mining. Now she is pursuing her Ph.d Computer Science in NGM College ,Bharathiar University. Her area of Interests Data mining, Knowledge Engineering and Image Processing.



Dr. Antony Selvadoss Thanamani is presently working as Professor and Head, Dept of Computer Science, NGM College, Coimbatore, India (affiliated to Bharathiar University, Coimbatore). He has published more than 100 papers in international/ national journals and conferences. He has authored many books on recent trends in Information Technology. His areas of interest include E-Learning, Knowledge Management, Data Mining, Networking, Parallel and Distributed Computing. He has to his credit 24 years of teaching and research experience. He is a senior member of International Association of Computer Science and Information Technology, Singapore and Active member of Computer Science Society of India, Computer Science Teachers Association, New York