

# A SURVEY ON OPINION MINING FROM ONLINE REVIEW SENTENCES

Dr.P.Perumal<sup>1</sup>,M.Kasthuri<sup>2</sup>

<sup>1</sup> Professor, Computer science and Engineering, Sri Ramakrishna Engineering College, TamilNadu, India

<sup>2</sup> ME Student, Computer science and Engineering, Sri Ramakrishna Engineering College, TamilNadu, India

\*\*\*

**Abstract** - Nowadays, if one wants to buy a product, consumer is no longer limited to asking friends and family for opinions because there are many user reviews and discussions in public forums on the Web about the product. Opinion targets from online reviews is an important and challenging task in opinion mining. Large amount of user generated data is present on web in the form of blogs, reviews tweets, comments etc. The internet users not only use the available resources in the web, but also give their feedback, thus generating additional useful information. Due to overwhelming amount of user's opinions, views, feedback and suggestions available through the web resources, it's very much essential to explore, analyze and organize their views for better decision making. Opinion Mining or Sentiment Analysis is a Natural Language Processing and Information Extraction task that identifies the user's views or opinions explained in the form of positive, negative or neutral comments and quotes underlying the text.

**Key Words:** Opinion Mining, Sentiment Analysis, Information Extraction, Natural Language Processing, Machine Learning, Lexicon Approach.

## 1. INTRODUCTION

Today's world is a world of Internet, almost all work can be done with the help of it, from simple mobile phone recharge to biggest business deals can be done with the help of this technology. People spent their most of the times on surfing on the Web; it becomes a new source of entertainment, education, communication, shopping etc. Users not only use these websites but also give their feedback and suggestions that will be useful for other users. In this way a large amount of reviews of users are collected on the Web that needs to be explored, analyze and organized for better decision making. Data mining (the analysis step of the knowledge discovery in databases process, or KDD) is the process of discovering new patterns from large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems.

The goal of data mining is to extract knowledge from a data set in a human-understandable structure and involves database and data management, data preprocessing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of found structure, visualization and online updating. The data mining process converts data into valuable knowledge that can be used for decision support and it is automated (no need for human intervention)[1].

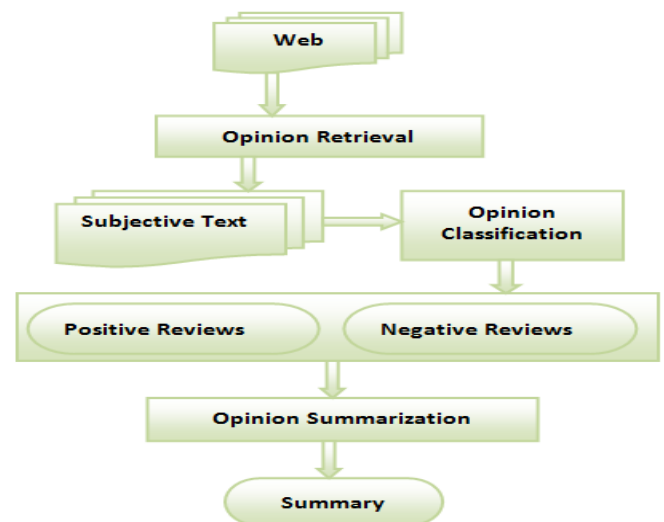


Figure 1: Architecture of Opinion Mining

In figure.1, the architecture of Opinion Mining which says how the input is being classified on various steps to summarize the reviews. The process of automatic extraction of knowledge by means of opinion of others about some particular product, topic or problem. Opinion mining is also called sentiment analysis due to large volume of opinion which is rich in web resources available online. Analyzing customer review is most important, it tend to rate the product and provide opinions for it which is been a challenging problem today.

Opinion feature extraction is a sub problem of opinion mining, with the vast majority of existing work done in the product review domain. Main fields of research in sentiment analysis are Subjectivity Detection, Sentiment Prediction, Aspect based Sentiment Summarization, Text summarization for opinions,

Contractive viewpoint, Summarization, Product Feature Extraction, Detecting opinion spam.

The idea of Opinion mining and Sentiment Analysis tool is to process a set of search results for a given item based on the quality and features. A key problem in this area is sentiment classification, where a document is labeled as a positive or negative evaluation of a target object (film, book, product etc...). In figure.2, depicts the major important steps in order to achieve an opinion impact. Workflow of Opinion Mining of how the opinions are being extracted from people review over their comment.

Sentiment analysis can be done at Document level, Sentence level, and Aspect or Feature level. In Document level the whole document is classified either into positive or negative class. Sentence level sentiment classification classifies sentence into positive, negative or neutral class. Aspect or Feature level sentiment classification concerns with identifying and extracting product feature from the source data.

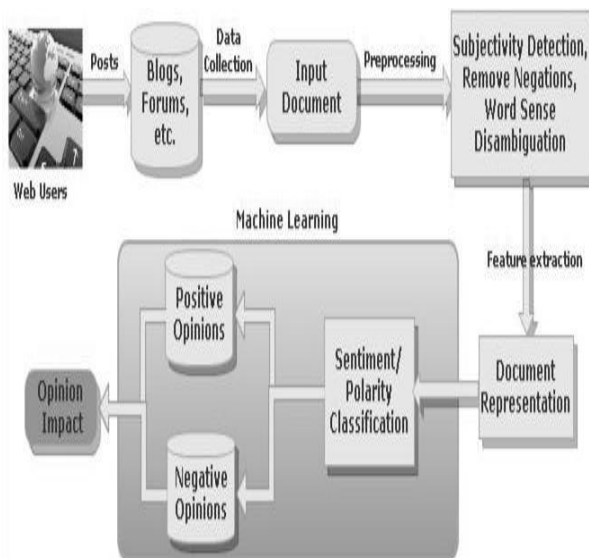


Figure 2: Systematic Workflow of Sentiment Analysis

The evaluation of opinion can be done in two ways:

**1. Direct opinion** gives positive or negative opinion about the object directly. For example, "The picture quality of this camera is poor" expresses a direct opinion.

**2. Comparison** means to compare the object with some other similar objects. For example, "The picture quality of camera-y is better than that of Camera-x." expresses a comparison.

## 1.1 OPINION MINING APPLICATIONS

Opinion mining and sentiment analysis cover a wide range of applications.

1. Argument mapping software helps organizing in a logical way these policy statements, by explicit the logical links between them. Under the research field of Online Deliberation, tools like Compendium, Cohere, Debate graph have been developed to give a logical structure to a number of policy statement, and to link arguments with the evidence to back it up.

2. Voting Advise Applications help voters understanding which political party (or other voters) have closer positions to theirs. For instance, SmartVote.ch asks the voter to declare its degree of agreement with a number of policy statements, then matches its position with the political parties.

3. Automated content analysis helps processing large amount of qualitative data. There are today on the market many tools that combine statistical algorithm with semantics and ontology, as well as machine learning with human supervision. These solutions are able to identify relevant comments and assign positive or negative connotations to it (the so-called sentiment).

## 1.2 CURRENT RESEARCH

Current research is focusing on:

- Improving the accuracy of algorithm for opinion detection
- Reduction of human effort needed to analyze content
- Semantic analysis through lexicon/corpus of words with known sentiment for sentiment classification
- Identification of policy opinionated material to be analyzed
- Computer-generated reference corpuses in political/governance field
- Visual mapping of bipolar opinion
- Identification of highly rated experts

## 2. RELATED WORK

Opinion target and opinion word extraction are not new tasks in opinion mining. There are many significant research efforts on opinion targets/words extraction (sentence level and corpus level). In sentence level extraction, previous methods mainly aimed to identify all opinion target/word mentions in sentences. They regarded it as a sequence labeling task, where several classical models were used, such as Conditional Random Fields(CRFs) and Support Vector Machine(SVM)[2]. This paper belongs to corpus level extraction, and aims to generate a sentiment lexicon and a target list rather than to identify mentions in sentences. Most of previous corpus-level methods adopted a co-

extraction framework, where opinion targets and opinion words reinforce each other according to their opinion relations.

In [3], proposed a technique based on association rule mining to extract product features. The main idea is that people often use the same words when they comment on the same product features. Then frequent item sets of nouns in reviews are likely to be product features while the infrequent ones are less likely to be product features. This work also introduced the idea of using opinion words to find additional (often infrequent) features.

Previous studies focused on opinion target extraction, can be divided into two main categories: supervised and unsupervised methods[4]. In supervised approaches, the opinion target extraction task was usually regarded as a sequence labeling task. The main limitation of these methods is that labeling training data for each domain is time consuming and impracticable. Product features and sentenced words are extracted using Double Propagation Algorithm. In unsupervised methods, most approaches regarded opinion words as the important indicators for opinion targets. They exploited an association rule mining algorithm and frequency information to extract frequent explicit product features in a bootstrapping process and designed some syntactic patterns to extract opinion targets. Compared to the traditional unsupervised alignment model, the proposed model obtains better precision because of the usage of partial supervision. In addition, when estimating candidate confidence, they penalized higher-degree vertices in our graph-based co-ranking algorithm to decrease the probability of error generation. Their experimental results on three corpora with different sizes and languages show that their approach effectively outperforms state-of-the-art methods.

In [5], this paper presents a method for identifying an opinion with its holder and topic, given a sentence in online news media texts. The semantic structure of sentence method uses semantic role labeling as an intermediate step to label an opinion holder and topic using FrameNet data. To identifies an opinion-bearing word, labels semantic roles related to the word in the sentence, and then finds a holder and a topic of the opinion word among labeled semantic roles. In this paper, they propose a novel method that employs Semantic Role Labeling, a task of identifying semantic roles given a sentence. They utilize FrameNet data by mapping target words to opinion-bearing words and mapping semantic roles to holders and topics, and then use them for system training.

Aspect based opinion mining is one of the level of Opinion mining that determines the aspect of the given reviews and classify the review for each feature[6]. In this paper an Aspect based Opinion Mining system named as "Aspect based Sentiment Orientation System" is proposed which extracts the feature and opinions from sentences

and determines whether the given sentences are positive, negative or neutral for each feature. Negation is also handled by the system. Dictionary based approach of the unsupervised technique is used to determine the orientation of sentences. To determine the opinion words and their synonyms and antonyms WordNet is used as a dictionary. The objective of this paper is to determine the polarity of the customer reviews of mobile phones at aspect level. System performs the aspect based opinion mining on the given reviews and the feature wise summarized results generated by the system will be helpful for the user in taking the decision.

Consumers are often forced to wade through many on-line reviews in order to make an informed product choice[7]. This paper introduces OPINE, an unsupervised information extraction system which mines reviews in order to build a model of important product features, their evaluation by reviewers, and their relative quality across products. OPINE is an unsupervised information extraction system which extracts fine-grained features, and associated opinions, from reviews. OPINE's use of the Web as a corpus helps identify product features with improved precision compared with previous work. OPINE uses a novel relaxation-labeling technique to determine the semantic orientation of potential opinion words in the context of the extracted product features and specific review sentences; this technique allows the system to identify customer opinions and their polarity with high precision and recall.

Opinion targets (targets for short) are entities and their attributes on which opinions have been expressed. To perform the tasks, there are several syntactic relations that link opinion words and targets. These relations can be identified using a dependency parser and then utilized to expand the initial opinion lexicon and to extract targets[8]. This proposed method is based on bootstrapping. In this paper focuses on two important tasks in opinion mining, i.e., opinion lexicon expansion and target extraction and a propagation approach to extract opinion words and targets iteratively given only a seed opinion lexicon of small size. 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 and also they propose novel methods for new opinion word polarity assignment and noisy target pruning.

Feature based opinion mining is one of the basic tasks in opinion mining is classifying the polarity of a given text or feature/aspect level to find out whether it is positive, negative or neutral[9]. Different methodologies are used for this purpose. Some expert analysts used the scaling system to associate numbers with appropriate sentiments that a word is depicting. The Product feature extraction is an important task of review mining and summarization. Opinion features are mined from product reviews based on data mining and natural language

processing methods. A feature-based summary of a large number of customer reviews of a product sold online is obtained. This problem will become increasingly important as more people are buying and expressing their opinions on the Web. This work tried to find out better product features from customer reviews.

In [10] this paper focuses on how to improve aspect-level opinion mining for online customer reviews. They propose a novel generative topic model, the Joint Aspect/Sentiment (JAS) model, to jointly extract aspects and aspect-dependent sentiment lexicons from online customer reviews. Then apply the extracted aspect dependent sentiment lexicons to a series of aspect-level opinion mining tasks, including implicit aspect identification, aspect-based extractive opinion summarization, and aspect-level sentiment classification. To consider more context information and more sources of knowledge to better identify opinion words and to incorporate more sources of signals, such as “and” rules in linguistics heuristics and synonym/antonym rules , to better identify aspect-aware sentiment polarities.

Review mining has recently received a lot of attention, which aims to discover the valuable information from the massive product reviews[11]. Product feature extraction is one of the basic tasks of product review mining. Its effectiveness can influence significantly the performance of subsequent jobs. A dependency grammar is employed to describe the relations between opinion words and product features. Based on these relations, the exaction rules are designed. Thus, opinion words can be recognized by identified features, and features can be identified by known opinion words. So the extracted opinion words and product features are utilized to identify new opinion words and new product features, which are used again to extract more opinion words and product features. Double Propagation is a state-of-the-art technique in product feature extraction. They apply the Double Propagation to the product feature exaction from Chinese product reviews and adopt some techniques to improve the precision and recall. In order to increase the recall, they introduce indirect dependency relations and verb product features.

Identifying an opinion target, a primary object of the opinion expression (e.g., the real-world object, event, and abstract entity), is helpful for extracting target-related opinions and detecting user interests[12]. This paper presents a novel framework for target-based opinion analysis, which extracts opinionated sentences and identifies their opinion targets from news articles. To presents two tasks, opinionated sentence judgment and opinion target identification, which are subtasks in NTCIR-8 MOAT. For the opinion judgment at the sentence-level, the opinion score and opinion patterns are used. If a sentence is matched to opinion patterns or contains strong or several opinion clue words, it would be detected as an opinionated sentence. For the opinion target

identification, they extract noun phrases as candidates and build a classifier, which determines whether a candidate is suitable for an opinion target or not, with four types of features: the appearance on the title section, document-level language model, opinion score of a candidate, and collocation information between a candidate and opinion clues.

Online customer reviews is considered as a significant informative resource which is useful for both potential customers and product manufacturers. In web pages, the reviews are written in natural language and are unstructured-free-texts scheme. The task of manually scanning through large amounts of review one by one is computational burden and is not practically implemented with respect to businesses and customer perspectives[13]. In this paper, a dependency and semantic based approach is proposed for mining opinions from online customer reviews. They focused on extracting relations between product features and opinions and proposed a novel way to capture the actual relations of product features in sentences regardless the distance from them to opinions. To develop ways to establish a correct relationship between the product feature (the topic of the sentiment) and the opinion word (the subjective expression of the product feature).

In [14], extracting sentiment and topic lexicons is important for opinion mining. A sentiment lexicon is a list of sentiment expressions, which are used to indicate sentiment polarity (e.g., positive or negative). The sentiment lexicon is domain dependent as users may use different sentiment words to express their opinion in different domains (e.g., different products). A topic lexicon is a list of topic expressions, on which the sentiment words are expressed. Extracting the topic lexicon from a specific domain is important because users not only care about the overall sentiment polarity of a review but also care about which aspects are mentioned in review. The co-extraction task of sentiment and topic lexicons in a target domain they do not have any labeled data, but have plenty of labeled data in a source domain. To leverage the knowledge extracted from the source domain to help lexicon co-extraction in the target domain. They propose a simple strategy to generate a few high-quality sentiment and topic seeds for the target domain and a novel Relational Adaptive bootstrapping (RAP) method to expand the seeds, which can exploit the relationships between topic and opinion words. The extracted sentiment lexicon can be applied to sentiment classification effectively.

They proposes a novel graph-based approach to extract opinion targets using word-based translation model (WTM)[15]. Given a collection of reviews , a WTM to identify potential opinion relations in all sentences, and then the associations between opinion targets and opinion words are estimated. Based on these associations, they exploit a graph-based algorithm to compute the

confidence of each opinion target candidate. Then the candidates with higher confidence scores are extracted as opinion targets. Compared with previous adjacent methods and syntax-based methods, by using WTM, they can capture opinion relations more precisely and therefore be more effective for opinion target extraction, especially for large informal Web corpora.

In [16] this paper proposes a novel monolingual word alignment (MWA) method to extract collocation of higher quality and with longer spans only from monolingual corpus, without using any additional resources. The monolingual corpus is replicated to generate a parallel corpus, where each sentence pair consists of two identical sentences in the same language, instead of a sentence in one language and its translation in another language. The bilingual word alignment algorithm to the monolingual scenario to align the potentially collocated word pairs in the monolingual sentences, with the constraint that a word is not allowed to be aligned with itself in a sentence. In addition, a ranking method was proposed to finally extract the collocations from the aligned word pairs. It scores collocation candidates by using alignment probabilities multiplied by a factor derived from the exponential function on the frequencies.

With the rapid expansion of e-commerce, people are more likely to express their opinions and hands-on experiences on products or services they have purchased. These reviews are important for both business organizations and personal costumers. Companies can decide on their strategies for marketing and products improvement. Customers can make a better decision when purchasing products or services. In [17] this paper, formulate the review mining task as a joint structure tagging problem. A new framework based on Conditional Random Fields is proposed. The framework can employ rich features to simultaneously extract object features, positive opinions and negative opinions. With this framework, to investigate the chain structure, conjunction structure and syntactic tree structure for review mining. A new unified model, called skip tree CRFs, is proposed for review mining. Through extensive experiments, we show that our proposed framework is effective. It outperforms many state-of-the-art methods.

Word alignment plays an important role in many NLP tasks as it indicates the correspondence between words in a parallel text. Although widely used to align large bilingual corpora, generative models are hard to extend to incorporate arbitrary useful linguistic information. Every sentence is replicated to generate a parallel corpus, and they apply the bilingual word alignment algorithm to the monolingual scenario to align a noun/noun phase with its modifier. Then plan to use other word alignment methods, such as discriminative model[18] for this task. Meanwhile, they add some syntactic information into WTM to constrain the word alignment process, in order to identify opinion relations

between words more precisely. To believe that there are some verbs or nouns can be opinion words and they may be helpful for opinion target extraction. And they think that it's useful to add some prior knowledge of opinion words (sentiment lexicon) in our model for estimating candidate opinion relevance.

In [19] this paper proposed a set of techniques for mining and summarizing product reviews based on data mining and natural language processing methods. The objective is to provide a feature-based summary of a large number of customer reviews of a product sold online. They indicate that the proposed techniques are very promising in performing their tasks. To believe that this problem will become increasingly important as more people are buying and expressing their opinions on the Web. Summarizing the reviews is not only useful to common shoppers, but also crucial to product manufacturers. For each discovered feature, related opinion sentences are put into positive and negative categories according to the opinion sentences' orientations. A count is computed to show how many reviews give positive/negative opinions to the feature. All features are ranked according to the frequency of their appearances in the reviews. Feature phrases appear before single word features as phrases normally are more interesting to users. Other types of rankings are also possible. For example, we can also rank features according to the number of reviews that express positive or negative opinions. To provide a feature-based summary of a large number of customer reviews of a product sold online. Let us use an example to illustrate a feature-based summary. Assume that, summarize the reviews of a particular digital camera, *digital\_camera\_1*.

The summary looks like the following:

*Digital\_camera\_1*:

Feature: **picture quality**

Positive: 253

<individual review sentences>

Negative: 6

<individual review sentences>

Feature: **size**

Positive: 134

<individual review sentences>

Negative: 10

<individual review sentences>

...

**Figure 3: An example summary**

In Figure 3, *picture quality* and (camera) *size* are the product features. There are 253 customer reviews that express positive opinions about the picture quality, and only 6 that express negative opinions. The <individual review sentences> link points to the specific sentences and/or the whole reviews that give positive or negative comments about the feature.

Syntax based methods usually exploited syntactic patterns to extract opinion targets[20], which were however prone to suffer from parsing errors when dealing with online informal texts. In contrast, alignment based methods used word alignment model to fulfill this task, which could avoid parsing errors without using parsing. To capture opinion relations in sentences by using syntactic patterns, employ the manual designed syntactic patterns proposed by [8], only the syntactic patterns based on the direct dependency are employed to guarantee the extraction qualities.

In [20], to use syntactic patterns and unsupervised word alignment model (WAM) to capture opinion relations. In addition, a partially supervised word alignment model (PSWAM) to incorporate syntactic information into WAM. To extract opinion targets from reviews, they adopt the framework proposed by [4], which is a graph-based extraction framework and has two main components. To capture opinion relations in sentences and estimate associations between opinion target candidates and potential opinion words and a potential opinion relation is comprised of an opinion target candidate and its corresponding modified word. To estimate the confidence of each candidate. The candidates with higher confidence scores than a threshold will be extracted as opinion targets. In this procedure, they formulate the associations between opinion target candidates and potential opinion words in a bipartite graph. A random walk based algorithm is employed and to estimate the confidence of each target candidate.

The performance variation of syntax based methods and alignment based methods on opinion target extraction task for the dataset with different sizes, different languages and different domains[20]. Through experimental results, choosing which method is not related with corpus domain and language, but strongly associated with the size of the corpus. They conclude that syntax-based method is likely to be more effective when the size of the corpus is small, and alignment-based methods are more useful for the medium size corpus.

### 3. CONCLUSIONS

This paper focuses on survey of opinion mining. It is not only encompasses concepts of text mining but also the concepts of information retrieval. Major challenges in opinion mining includes feature weighting which plays a crucial role for good classification. Various features extraction approaches are listed this survey. To enhance

the customer satisfaction, merchants and product manufacturers allow customers to review or express their opinions on the products or services. To captures opinion relations more precisely and therefore is more effective for opinion target and opinion word extraction. Then each candidate will be assigned a confidence and ranked, and the candidates with higher confidence than a threshold will be extracted as the results.

In future, Opinion Mining can be carried out on a set of reviews and set of discovered feature expressions extracted from reviews. The state-of-art for current methods, useful for producing better summary based on feature based opinions as positive, negative or neutral is the Expectation Maximization algorithm based on Naïve Bayesian is the most efficient method.

### REFERENCES

- [1] G.Angulakshmi, Dr.R.ManickaChezian, "An Analysis on Opinion Mining: techniques and tools", Vol. 3, Issue 7, July 2014
- [2] Kang Liu, Liheng Xu and Jun Zhao, "Extracting Opinion targets and Opinion Words from Online Reviews with Graph Co-ranking", Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pages:314-324, Baltimore, Maryland, USA, June 2014.
- [3] Lei Zhang, Bing Liu, Suk Hwan Lim and Eamonn O'Brien-Strain, "Extracting and Ranking Product Features in Opinion Documents", Coling 2010: Poster Volume, pages 1462-1470, Beijing, August 2010.
- [4] Kang Liu, Liheng Xu, Yang Liu and Jun Zhao, "Opinion Target Extraction Using Partially-Supervised Word Alignment Model", Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, 2009.
- [5] Soo-Min Kim and Eduard Hovy, "Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text", 2009.
- [6] Richa Sharma<sup>1</sup>, Shweta Nigam<sup>2</sup> and Rekha Jain<sup>3</sup>, "Mining of product reviews at aspect level", Vol.4, No.3, May 2014.
- [7] Ana-Maria Popescu and Oren Etzioni, "Extracting product features and opinions from reviews", 2005.
- [8] Guang Qiu, Bing Liu, Jiajun Bu and Chun Chen, "Opinion Word Expansion and Target Extraction through Double Propagation", vol.1, 2006.
- [9] G. vinodhini, L. srisubha and RM. chandrasekaran, "feature based opinion mining for customer reviews", 2012.
- [10] Xu Xueke, Cheng Xueqi, Tan Songbo, Liu Yue, and Shen Huawei, "aspect-level opinion mining of online customer reviews", 2013.
- [11] Yahui xi, "Extracting Product Features from Chinese Product Reviews", journal of multimedia, vol. 8, no. 6, december 2013.

- [12] Yoonjung Choi, Seongchan Kim and Sung-Hyon Myaeng, "Detecting Opinions and their Opinion Targets in NTCIR-8",2010.
- [13] Gangarn Somprasertsri and Pattarachai Lalitrojwong , "Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization", Journal of Universal Computer Science, vol. 16, no. 6 (2010), 938-955.
- [14] Fangtao Li, Sinno Jialin Pan, Ou Jin, Qiang Yang and Xiaoyan Zhu, "Cross-Domain Co-Extraction of Sentiment and Topic Lexicons",2012.
- [15] Kang Liu, Liheng Xu and Jun Zhao, "Opinion Target Extraction Using Word-Based Translation Model",2010.
- [16] Zhanyi Liu, Haifeng Wang, Hua Wu and Sheng Li.2009."Collocation Extraction Using Monolingual Word Alignment Model". In *Proceedings of EMNLP,2009*.
- [17] Fangtao Li, Chao Han, Minlie Huang, Xiaoyan Zhu, Ying-Ju Xia, Shu Zhang and Hao Yu, " Structure-Aware Review Mining and Summarization", Proceedings of the 23rd International Conference on Computational Linguistics, pages 653–661,Beijing, August 2010.
- [18] Yang Liu, Qun Liu, and Shouxun Lin.2010,"Discriminative word alignment by linear modeling", Computational Linguistics, 36(3):303–339.
- [19] Minqing Hu and Bing Liu, "Mining and Summarizing Customer Reviews", *KDD'04*, August 22–25, 2004, Seattle, Washington, USA .Copyright 2004 ACM.
- [20] Kang Liu, Liheng Xu and Jun Zhao, "Syntactic Patterns versus Word Alignment: Extracting Opinion Targets from Online Reviews", Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 1754–1763,2013.