

# A Study of Different Approaches to Aspect-based Opinion Mining

Pratima More  
Department of Computer Engg,  
PICT, Pune

Archana Ghotkar  
Department of Computer Engg,  
PICT, Pune

## ABSTRACT

In recent years, Opinion mining has been an active research area in Text mining and analysis, natural language processing. Opinion mining is the computational study of people's opinion expressed in written language or text towards entities and their aspects. With the growth of internet, social networking sites, blogs, discussion forums, e-commerce websites have gained a tremendous importance and have provided platform for people to express and share their opinion on entities and their aspects. As opinionated web content is increasing rapidly in the form of reviews, comments, blogs, status updates, tweets, etc. it is practically impossible for people or organization to analyze all opinions at a time to make good decisions. Hence, there is a need for effective automated system to evaluate opinions and generate accurate results. This paper describes opinion mining and focuses on the sub topic aspect-based opinion mining, tasks in aspect-based opinion mining, current state-of-the-art methods used for aspect-based opinion mining, advantages and disadvantages of these methods and latest research challenges in aspect-based opinion mining. Our experimental results based on some of the aspect extraction techniques, gives an idea of which aspect extraction techniques are efficient and yield accurate results in practical opinion mining applications.

## General Terms

Data Mining, Text Mining, Natural Language Processing

## Keywords

Opinion Mining, Text Analysis, aspect-based opinion mining, aspect extraction, opinion polarity detection.

## 1. INTRODUCTION

Content growth on the Internet in recent years has made a huge volume of opinionated data available. Especially in the automotive, electronics gadgets, e-commerce, restaurants and movie sectors, customers write reviews about products, service or their features. By analyzing these reviews, new customers find others opinion and experience about different features of the product or service. They can compare the products to each other to find the best one that meet their needs. By analyzing reviews, manufacturers/firms/organization can find out strengths and weaknesses of their products/service or those of their competitors. In this way, they can solve the reported problems, improve the product/service based on feature request or complaints of people in reviews and thus maintain good customer relations needed to boost the business. As such opinionated web content is very large, manual analysis of reviews by people and organization for decision making is not possible and time consuming. Taking all these scenarios into consideration there is a need

of automating the process of analyzing large data and extracting opinions.

The automated process of analyzing opinionated data in form of written language or text, to find out people's opinion is called opinion mining. Opinion Mining is studied at three levels, document level, sentence level and aspect/phrase level. In document level, the entire document is input and task is to find overall opinion in document. Few drawbacks of document level are: 1) Documents may contain unrelated sentences to target entity, hence processing such sentences is time consuming. 2) Document may be describing aspects of multiple product, so finding multiple aspect of multiple product and their opinion polarity from the document is complex. Sentence level is a two-step process. First, classify the sentence as subjective or objective. Second, classify opinion polarity of each subjective sentence. In sentence level analysis, sentence may be expressing opinion about multiple aspects of different products, so identification of these and their opinion orientation is difficult.

Aspect level carries out fine-grained opinion analysis, since it works at word/phrase level. Aspects of entity are extracted and opinion polarity of each extracted aspect is calculated. This leads to a new direction called aspect-based opinion mining, also known as feature-based opinion mining. For example, "Camera quality of iPhone is good but battery life is poor" finds two aspects (Camera quality, Battery) of iPhone. The opinion polarity on camera quality aspect is positive and battery is negative. Aspect level analysis, helps to find out multiple aspects of different product from the reviews.

## 2. CORE TASKS IN ASPECT-BASED OPINION MINING

The major tasks in aspect-based opinion mining are aspect extraction and aspect opinion classification.

### 2.1 Aspect Extraction

This task involves identification of aspects, features of a target entity from the review. For example, "iPhone camera quality is amazing", the task of aspect extraction will find aspect as "camera quality" of entity "iPhone".

### 2.2 Aspect Opinion Classification

This task involves classifying extracted opinion words related to aspects into one of polarity scales (positive, negative, neutral etc.). In the example, "iPhone camera quality is amazing", the aspect opinion classification task will find opinion orientation of opinion word "amazing" on the "camera quality" aspect as positive.

### 3. DIFFERENT APPROACHES TO ASPECT EXTRACTION

Aspect extraction means information extraction task. Aspect is defined as feature of entity on which user comments in review. For example, in a mobile phone review, its aspect would be “camera quality”, “battery life”, “screen” etc. Researchers have used various approaches to solve this problem of aspect extraction. Figure 1 depicts some approaches used for aspect extraction in aspect-based opinion mining.

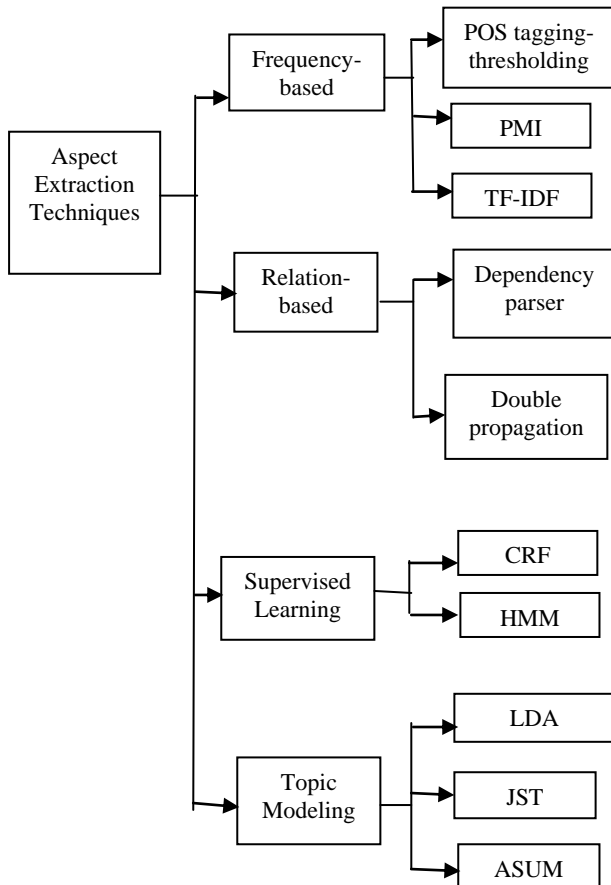


Figure 1: Different approaches to Aspect Extraction.

#### 3.1 Frequency-based Approach

This technique involves finding frequent nouns and noun phrases from the review as aspects. This method finds explicit aspect that are nouns and noun phrases from reviews in a given domain. This approach works, as while commenting on aspects of an entity, the vocabulary that people use is similar and converges [2].

In Hu and Liu [2], part-of-speech (POS) tagger is used to identify noun and noun phrases. Occurrence frequency of these noun and noun phrases is counted. A threshold is decided by manual tuning and only frequent noun and noun phrases whose count is greater than threshold is considered.

Popescu et al [3] improved the accuracy by removing noun phrases that are not aspects. It computes a pointwise mutual information score between the phrase and meronymy discriminators related with the entity class, e.g., a mobile phone class. The meronymy discriminators for the mobile class are, “of mobile”, “mobile has”, “mobile comes with”, etc.

In Blair-Goldensohn et al. [4], frequency based approach is refined by taking into consideration only those noun phrases that are in opinion-containing sentence. Ku et al. [5] used TF-IDF scheme to find frequent terms in reviews that are major topics. Moghaddam and Ester [6] improved frequency-based approach, by removing non-aspect terms with help of syntactic pattern-based filter.

In Scaffidi et al. [7], to identify true aspects, frequency of extracted noun and noun phrases from reviews is compared with their occurrence rates in English corpus.

Zhu et al. [8] uses a method based on the Cvalue measure for extracting multi-word aspects.

#### 3.2 Relation-based Approach

The idea behind this approach is to find relationship between aspect and opinion words to identify aspects. This approach uses grammatical relation and syntactic patterns between aspect and opinion words to find extraction rules.

In Zhuang et al. [10], a dependency parser is used to identify dependency relation for aspect extraction in reviews. After parsing, linkage of words in a sentence is done by a certain dependency relation like <NN-noun, JJ-adjective>, <NP-noun phrase, JJ-adjective>.

In Wu et al. [11], phrase dependency parser is used to extract noun phrases and verb phrases as aspect candidates. Dependency parser helps to identify dependency of individual words, but phrase dependency parser helps identify dependency of phrases, which is more accurate for aspect extraction.

Double propagation method is used in Qiu et al [12], in which different relations between opinion words and aspect, opinion words and aspects themselves, is used to design extraction rules.

#### 3.3 Supervised Learning Approach

Supervised learning approach involves inferring a model, from training data to apply for unlabeled data. Thus, identifying aspects, opinion and their polarity can be seen as a labelling problem where patterns and syntactic dependency is learned from labelled data and applied to unlabeled data. Hidden Markov Models (HMM) and Conditional Random Fields (CRF) are supervised learning methods used to extract aspects.

Jin and Ho [13] used lexicalized HMM model to learn patterns to extract aspects and opinion expressions.

Jakob and Gurevych [14] uses Conditional Random Fields (CRF) to train review sentences from different domains for domain independent extraction.

Li et al [15] uses variation of CRF, i.e., Skip-CRF and Tree-CRF, to find aspects and opinions.

#### 3.4 Topic Modeling Approach

Topic modeling is an unsupervised learning method for discovering topics from text document assuming that documents consists of mixture of topics and each topic is probability distribution over words. Two basic models used are pLSA (Probabilistic Latent Semantic Analysis) and LDA (Latent Dirichlet allocation) (Blei et al [17]). In context of opinion mining, discovered topics from topic models are aspects. Hence, Topic modeling can be used for aspect extraction. Topic modeling help in aspect grouping and cover both aspect and opinion words.

Zhao et al. [18] proposed MaxEnt-LDA (Maximum Entropy and LDA combination) hybrid model to discover both aspect words and aspect-specific opinion words jointly.

Chenghua Lin et al. [20] proposed Joint Sentiment Topic (JST) model. JST model considered topics and sentiments together. JST model focuses on the extraction of opinion-aspect pair.

Samuel Brody et al. [19] used local LDA model to find aspects. After aspect detection, for opinion word, adjectives are selected.

Yohan Jo et al. [21] concentrates on automated identification of aspects. This paper proposed Sentence-LDA (SLDA) model that assumes, words from a single sentence belongs to one aspect. For identification of opinion of different model Sentence-LDA is extended to Aspect and Sentiment Unification Model (ASUM). The ASUM model finds result in form of aspect, opinion pair.

Xu Xueke et al. [22] proposed Joint Aspect/Sentiment Model (JAS). JAS model use LDA [17] to extract aspect and aspect related terms.

#### 4. DIFFERENT APPROACHES TO ASPECT OPINION CLASSIFICATION

Aspect opinion classification task involves determining the opinion orientation expressed on each aspect. There are two main approaches to aspect opinion classification:

##### 4.1 Supervised Learning Approach

In this approach, classification techniques like Naive Bayes Classifier [24] [23], Support Vector Machine (SVM) [24], Decision Tree classifier [24], kNN classifier [25] are used for

Table 1. Strength and Limitation of aspect extraction techniques.

Aspect extraction approach	Strength	Limitations
Frequency-based method	Simple and quite effective method.	1) It generates many non-aspects and miss out low-frequency aspects.  2) Needs manual tuning of number of parameters, making difficult to port to another dataset.
Relation-based method	Helps find low frequency aspect.	It may produce many non-aspects which match the pre-defined syntactic pattern.
Supervised learning method	It overcomes frequency and relation-based method limitation by learning model	This technique needs manually labelled data for training. Hence, accuracy of learned model

	parameters from the training dataset.	depends on how accurately the training data is labelled for aspects and non-aspects.
Topic modeling method	1) No need of manually labelled data.  2) It performs aspect extraction and grouping task simultaneously in unsupervised manner.	It requires a large volume of data.

text categorization and classifying opinion words on aspects into one of the polarity scales. Problem with supervised learning techniques is, they are dependent on the training data. Hence, a classifier trained from labelled data in one domain often performs poorly in another domain.

##### 4.2 Lexicon-based Approach

It is unsupervised technique. The lexicon approach finds opinion orientation of review text using SentiWordNet. Work in [26], [27] proposes the use of SentiWordNet for classification of opinion words in opinion mining. Lexicon based method performs quite well in a large number of domains.

#### 5. RESEARCH CHALLENGES IN ASPECT-BASED OPINION MINING

##### 5.1 Implicit Aspect Finding

Finding implicit aspects is a challenging task. For example, “This mobile will not fit in a pocket”, here 'Fit in pocket' tells us about 'size' aspect of mobile. Research work needs to be done for developing techniques for extraction of such implicit aspects.

##### 5.2 Multiple Aspect Finding

Many reviews refer to multiple aspects of different products. Consider this example: “Camera quality of iPhone is good but battery life is poor”. This review refers to two aspects of iPhone i.e. Camera and battery. It is required to resolve these aspects and consider their polarity separately since user has positive opinion about camera quality and negative opinion about battery. Multiple aspect finding is thus a challenging task.

##### 5.3 Cross Domain Adaptation

Most of the opinion mining system is domain dependent. The same opinion word may indicate different polarity in different domain. It's crucial to have domain knowledge before mining opinion from the reviews. Consider this example, “Please go and read the book” tells that its positive opinion in a book review and negative opinion in case of movie review. For researchers it is challenging task to develop domain-independent methods and algorithms.

## 6. DISCUSSION

Based on the study done of various aspect extraction techniques, strengths and limitations of each of these methods is discussed in Table 1.

## 7. EXPERIMENTAL RESULTS

In this paper, frequency based method and relation based method is implemented for aspect extraction task. To evaluate results, the dataset used is Dropbox mobile app reviews. Figure 2 shows the precision and recall values for frequency based and relation based method for the task of aspect extraction from Dropbox app review dataset. Figure 3 shows F-score measure for aspect extraction task on Dropbox review using frequency based and relation based method.

The F-score measure for frequency based method is 38% and for relation based method it's 52%, thus aspect extraction task accuracy is increased with relation based approach. The experimental results show that relation based method yields on an average 10% more accurate results as compared to frequency based method. Thus it can be said that relation based method is efficient and can be used in practical opinion mining applications.

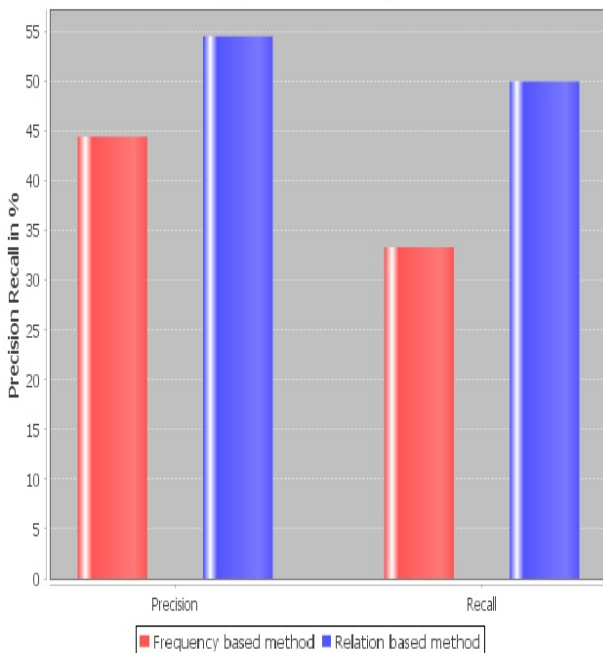


Figure 2: Precision and recall values for aspect extraction techniques.

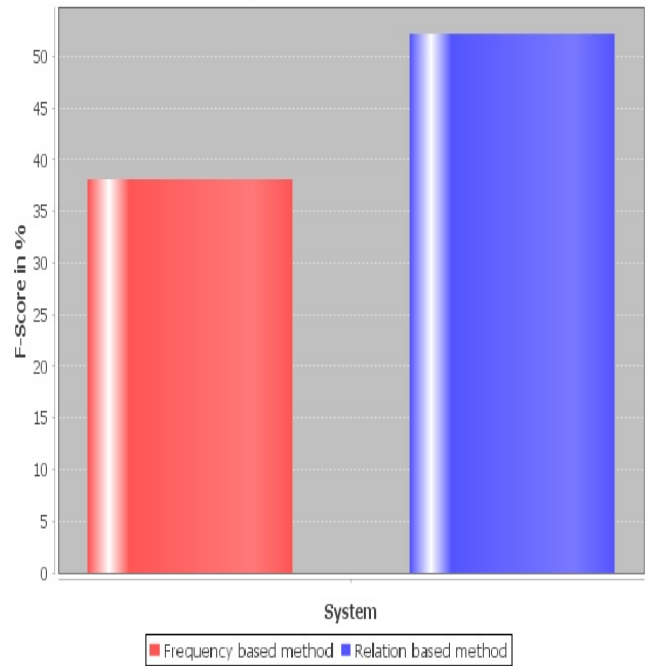


Figure 3: F-score value for aspect extraction techniques.

## 8. CONCLUSION

With the explosive growth of social media on the Web, organizations are increasingly relying on opinion mining methods to analyze the content of these media for their decision making. Aspect-based opinion mining, which aims to obtain detailed information about opinions, has attracted a great deal of attention from both the research community and industry. It has been a very active area of research in recent years due to many challenging research problems and also involves more practical applications in various fields. This paper has given a brief introduction of opinion mining, sub topic aspect based opinion mining, core tasks in aspect-based opinion mining. In this paper, a detailed study of different approaches for aspect extraction and aspect opinion classification, their strength and limitations and latest research challenges in aspect-based opinion mining is done. The experimental results based on some of the aspect extraction techniques show that relation based method yields on an average 10% more accurate results as compared to frequency based method.

## 9. REFERENCES

- [1] Liu, Bing, "Sentiment Analysis and Opinion Mining", Synthesis Lectures on Human Language Technologies, Morgan and Claypool Publishers, 2012.
- [2] Hu, Minqing and Bing Liu, "Mining and summarizing customer reviews", In Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), 2004
- [3] Popescu, Ana-Maria and Oren Etzioni, "Extracting product features and opinions from reviews", In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2005), 2005
- [4] Blair-Goldensohn, Sasha, Kerry Hannan, Ryan McDonald, Tyler Neylon, George A. Reis, and Jeff Reynar, "Building a sentiment summarizer for local service reviews", In Proceedings of WWW-2008 workshop on NLP in the Information Explosion Era, 2008

- [5] Ku, Lun-Wei, Yu-Ting Liang, and Hsin-Hsi Chen, "Opinion extraction, summarization and tracking in news and blog corpora", In Proceedings of AAAI-CAAW'06, 2006
- [6] Moghaddam, Samaneh and Martin Ester, "Opinion digger: an unsupervised opinion miner from unstructured product reviews", In Proceeding of the ACM Conference on Information and Knowledge Management (CIKM-2010), 2010.
- [7] Scaffidi, Christopher, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng, and Chun Jin, "Red Opal: product-feature scoring from reviews", In Proceedings of Twelfth ACM Conference on Electronic Commerce (EC-2007), 2007.
- [8] Chunliang Zhang, Jingbo Zhu, "Multi-class bootstrapping learning aspect-related terms for aspect identification", International Conference on Natural Language Processing and Knowledge Engineering, NLP-KE 2009, pp. 1-6, 24-27 Sept. 2009.
- [9] Varghese R., Jayasree M, "Aspect based Sentiment Analysis using support vector machine classifier", International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1581-1586, 2013.
- [10] Zhuang, Li, Feng Jing, and Xiaoyan Zhu, "Movie review mining and summarization", In Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2006), 2006
- [11] Wu, Yuanbin, Qi Zhang, Xuanjing Huang, and Lide Wu, "Phrase dependency parsing for opinion mining", In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2009).
- [12] Qiu, Guang, Bing Liu, Jiajun Bu, and Chun Chen, "Opinion Word Expansion and Target Extraction through Double Propagation", Journal Computational Linguistics, Volume 37 Issue 1, March 2011, Pages 9-27.
- [13] Jin, Wei and Hung Hay Ho, "A novel lexicalized HMM-based learning framework for web opinion mining", In Proceedings of International Conference on Machine Learning (ICML-2009).
- [14] Jakob, Niklas and Iryna Gurevych, "Extracting Opinion Targets in a Single-and Cross-Domain Setting with Conditional Random Fields", In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2010), 2010.
- [15] Li, Binyang, Lanjun Zhou, Shi Feng, and Kam-Fai Wong, "A Unified Graph Model for Sentence-Based Opinion Retrieval", In Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010).
- [16] Bin Lu, Myle Ott, Claire Cardie, Benjamin K. Tsou, "Multi-aspect Sentiment Analysis with Topic Models", In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, ICDMW '11, Pages 81-88, 2011.
- [17] David Blei, Andrew Ng and Michael Jordan, "Latent Dirichlet allocation", The Journal of Machine Learning Research, 2003, Volume 3, Pages 993-1022.
- [18] Zhao, Wayne Xin, Jing Jiang, Hongfei Yan, and Xiaoming Li, "Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid", In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2010).
- [19] Samuel Brody, Noemie Elhadad, "An unsupervised aspect-sentiment model for online reviews", In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT '10), Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 804-812, 2010.
- [20] Chenghua Lin, Yulan He, "Joint sentiment/topic model for sentiment analysis", In Proceedings of the 18th ACM conference on Information and knowledge management (CIKM '09), ACM, New York, NY, USA, pp. 375-384, 2009.
- [21] Yohan Jo, Alice H. Oh, "Aspect and sentiment unification model for online review analysis", In Proceedings of the fourth ACM international conference on Web search and data mining (WSDM '11), ACM, New York, NY, USA, pp. 815-824, 2011.
- [22] Xu Xueke, Cheng Xueqi, Tan Songbo, Liu Yue, Shen Huawei, "Aspect-level opinion mining of online customer reviews", Communications, China, vol.10, Issue 3, pp. 25-41, 2013.
- [23] N. Anitha, B. Anitha, S. Pradeepa, "Sentiment Classification Approaches –A Review", International Journal of Innovations in Engineering and Technology (IJET), Vol. 3 Issue 1, October 2013.
- [24] Vivek Narayanan, Ishan Arora and Arjun Bhatia, "Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model", 14th International Conference, IDEAL 2013, Hefei, China, October 20-23, 2013, pp 194-201.
- [25] Heui Lim, "Improving kNN Based Text Classification with Well Estimated Parameters", LNCS, Vol. 3316, Oct 2004, Pages 516-523.
- [26] Esuli, A, Sebastiani, F, "SentiWordNet: A publicly available resource for opinion mining", In Proceedings of the 6th international conference on Language Resources and Evaluation (LREC'06), 2006, pp.417-422.
- [27] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani, "Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining", In Proceedings of the 10th International conference on Language Resources and Evaluation (LREC'10), 2010.