

# A Survey on Aspect-based Opinion Polling from Online (web) Reviews

N. S. Ambekar<sup>1</sup> and Prof .N. L. Bhale<sup>2</sup>

<sup>1</sup>Department of Computer Engineering, MCOERC, University of Pune, Nashik, India.

<sup>2</sup>Department of Information Technology, MCOERC, University of Pune, Nashik, India.

**Abstract-**World Wide Web has almost changed the picture of shopping, trading etc in e-commerce. Opinions expressed by customers for any products or services in the form of reviews or ratings play a crucial role in making decisions. Conventional (Structured) methods were very hectic and expensive as they needed the total involvement of customers for answering the specially designed questions. Sentiment analysis hence plays here a vital role to express the opinions in the form of pairs i.e. (feature, polarity). Many techniques have been proposed to express these opinions in terms of polarity till date .Hereby, we present a survey on some current approaches implemented for opinion poll generation.

**Keywords:** *Opinion polling, Sentiment analysis, customer reviews*

## I. INTRODUCTION

Internet is an important platform for expressing ideas or views in the form of reviews, posts, blogs etc. Opinion mining (Sentiment Analysis) is a computational study of opinions, sentiments, subjectivity, evaluations, attitudes, appraisal, affects, views, emotions, etc., expressed in text. Sentiment analysis is more widely used in industry as well as in academia. "Opinions" are key influencers of our behaviors as whenever we need to make a decision, we often find out the opinions of others. In the past, Individuals seek opinions from friends and family while Organizations use surveys, focus groups, opinion polls, consultants. Organizations spend a huge amount of money to find consumer opinions using consultants, surveys and focus groups, etc. Many websites such as Amazon, IDMB etc provoke customer to post reviews on products, features, services etc. In order to enhance business growth, customer satisfaction in the form of these reviews is discovered which was either done traditionally by designing questions that are to be answered or now-a-days as free-form, on the spot customer feedback. With the increase in the number of people visiting the sites, hundreds and thousands of reviews are generated which creates a problem for customers to read them to buy a product at the same time

the companies to keep a track of the customer opinions or sentiments on their products and services. Such structured survey method was useful for only small data but with the explosion of social media structured survey technique suffered with the drawbacks like expenses in designing questions, lack of participation of customers. Hence, in this paper we focus on sentiment analysis based on unstructured spontaneous freeform customer reviews. The uneven and dynamic nature of the web has lead to the research in the field of opinion mining or sentiment analysis. Many recent studies have put forward sentiment-based or opinion-oriented summarization [5], [6], [13], [8]. The basic approach of sentiment analysis is to collect, analyze and extract sentiments or opinions from the customer reviews and produce the summarized results in the form of an opinion poll. an opinion poll. Many a times, opinions are hidden in long forum posts and blogs. It becomes very difficult for customers to find relevant information of their interest, extract, summarize and organize the same to be used. Automated classification of sentiments have become an important research area that help companies and organizations to find out reviews for their product's purchased (for ex. Mobiles sold) or services (for ex. Movies).The results are in the form of opinions such as positive ,negative, or neutral opinions on products or business. The major steps carried out for opinion mining are(1) to discover the features of the products pointed out in the reviews and(2) to find whether the comments are positive or negative. Many a times, now people express their positive and negative sentiments clearly in the form of ratings, which can be easily converted into an opinion poll. However, nowadays people mostly express their opinions in free-form textual reviews without assigning any ratings. Supervised document classification algorithms were used in some previous studies [12], [19], [16] to analyze customer reviews in text form. However, this approach suffered with the drawbacks (1) It required labeled training data which was costly and time consuming,(2) Opinion poll generated may not be in proper meaningful format. For ex, to analyze a customer review on a mobile" Mobile battery

life is not so good". Here the review expresses a negative opinion on the battery aspect. Therefore, aspect-based opinion analysis techniques [18] are needed.

Previous work feature based sentiment summarization [5], [6] that addressed aspect-based opinion polling is implemented on the sentence level instead of the document level, and does not address the polarity conflicting issue of document-level aspect-based opinion analysis. Implicit aspect expression problem, multiword terms problem couldn't be solved by this technique. Customers may express different opinions on multiple aspects in the single review statement. For example, for a mobile review "Mobile is slim and light weight but it has short battery life" aspects i.e. positive body aspect as well as negative memory aspect. To deal with such multiple aspect sentences and segment them into single multiple aspects, aspect opinion polling method is used [22]. This approach worked at sub-sentence level but failed at word level and seeds were manually selected which is troublesome. For example, for a mobile review "Mobile has very attractive color and body but it has less memory storage" contains two aspects i.e. positive body aspect as well as negative memory aspect.

This paper is organized as follows: Section 2- discusses related work in this field, Section 3- reveals the basic structure for aspect-based opinion poll generation, Section 4- Different approaches used for opinion polling, Section 5-Parameters required to evaluate the performance.

## II. LITERATURE SURVEY

Most of the research work has been carried out for opinion mining or sentiment analysis for summarizing the sentiments from customer reviews [13], [18], [9], [14], [5], [6]. Recent research work focused upon product reviews for aspect/feature based opinion analysis. Segmentation of text is an issue in retrieval of information. Previous studies had focused on text segmentation at document level instead of the sentence level. Feature-based classification and sentence summarization [5], [6] revealed the results in form of sentiment summary as per the feature of the product. B. Pang, Lee et.al [18] proposed a technique to analyse textual reviews to predict polarities using supervised document classification algorithms. M. Hu and B.Liu [7] proposed a set of techniques for mining and summarizing product reviews based on data mining and natural language processing methods. They focused only on explicit features but couldn't deal with implicit aspect identification. Stemming and fuzzy matching techniques used by them failed to deal the multiple aspect sentences and document-level polarity conflict problems. B.Pang and L. Lee, [10]

focused on supervised or semi-supervised learning techniques for sentiment analysis which need labelled data for training. Kim and Hovy [8] presented a technique for finding out opinion topics based on semantic frames, but presented a limited evaluation. But, their work did not discuss the issue of segmentation of multi-aspect sentences for aspect mention extraction or opinion topic extraction. Some research scholars [19], [23], [24] have implemented single aspect bootstrapping technique for learning subjective words or sentiment patterns.

Thomas et al. [36] proposed a supervised SVM-based classifier to resolve the problem of the support/oppose sentiment classification problem. In some of the recent work [24], [25], a ranking or ordinal regression framework was used for polarity analysis based on a multipoint scale and stressed on supervised or semi-supervised learning techniques for sentiment analysis which required labelled data for training. Text segmentation is an important problem in information retrieval. Previous studies [2], [27], [28] focused on text segmentation at the document level instead of the sentence level. Jingbo Zhu [22] performed comparison experiments, by evaluating a state-of-the-art linear text segmentation technique for aspect-based sentence segmentation. Results showed that the linear text segmentation technique yielded unsatisfactory performance in the aspect-based sentence segmentation task because a single sentence cannot provide sufficient context to determine topic changes for segmentation.

Titov and McDonald [18] proposed a statistical model for sentiment analysis based on the labelled customer reviews; customers manually entered the ratings for different aspects. J. Zhu, H. Wang, et.al, [21] in their work titled Multi-Aspect Opinion Polling from Textual Reviews, presented an unsupervised approach to aspect-based opinion polling from raw textual reviews without explicit ratings. Jingbo Zhu, et.al, [22] proposed opinion polling method that does not require labelled training data. They proposed an MAS model to segment a multi-aspect sentence into multiple single-aspect units for aspect-based opinion polling. They carried out analysis on a real Chinese restaurant reviews revealing 77 percent accuracy in aspect-based opinion polling tasks. This method is easy to implement and are applicable to other domains like product or movie reviews.

## III. DESIGN OF ASPECT-BASED OPINION POLL GENERATION MODEL

An aspect-based opinion poll generation process includes the following steps:

1. *Customer review database generation*-The customer reviews from different review sites, blogs are collected and product aspects commented by the customer in the review

are extracted. Web crawlers also known as spiders or robots can be used to download reviews by visiting different sites that are analyzed further. Natural language processing and Data mining techniques are used for mining.

2. *Pre-processing of reviews*- In preprocessing only the meaningful words are extracted from the free from unlabelled review database. The unwanted ones are removed.

3. *Parts-of-speech (POS) Tagging*- The product aspects and opinion expressing words are separated using POS tagger such as Stanford POS Tagger. The opinion words are generally adjectives and the product aspects are nouns. For ex:” Mobile color is attractive”, here mobile color is aspect (noun) and attractive is adjective.

#### 4. *Aspect Extraction*-

In aspect extraction, product aspects are extracted from each review sentence which is nouns. In review, features may be mentioned explicitly or implicitly by the reviewer. Features which are mentioned in a sentence directly are called as explicit features and features which are not mentioned directly are called implicit features. For example,

“Memory space of mobile phone is less”

In this sentence reviewer has discussed memory space directly so it is explicit feature and hence easy to extract.

Whereas in following sentence,

“This mobile phone stores fewer messages”

In this sentence reviewer is discussing about memory space of mobile phone but it is not conveyed directly in the sentence. So here memory space is implicit aspect. These aspects are difficult to understand and extract from sentence.

#### 5. *Extraction of Opinion Words and polarity identification Extraction*

The opinions in the review are identified which are generally adjectives and the opinions are classified as positive, negative or neutral to reveal the semantic orientation of the opinion words. WordNet can be used to identify the semantic orientation and the opinion orientation of each sentence is decided.

#### 6. *Final precise opinion poll Generation*

Final output in form of opinion poll is generated as a result of previous steps carried out above. Opinion poll generated can disclose the results in form of tables and graphs.

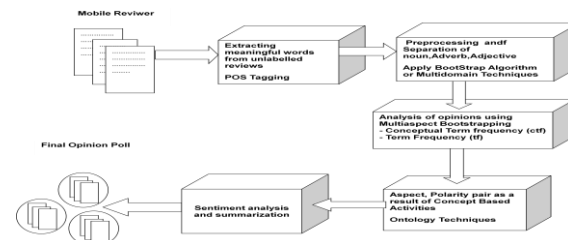


Figure 1 Block Diagram of the Opinion poll generating system

### IV. DIFFERENT APPROACHES USED FOR OPINION POLLING

Many approaches are being used for opinion polling, some are mentioned below.

#### A. *Lexicon-Based Approach*

Xiaowen Ding, Bing Liu, Philip S. Yu [29]proposed holistic lexicon-based approach which is an effective method for identifying semantic orientations of opinions expressed by reviewers on product aspects. It is able to deal with two major problems with the existing methods, (1) opinion words whose semantic orientations are context dependent, and (2) aggregating multiple opinion words in the same sentence. Opinion word lexicons can be found through bootstrapping process using WordNet. In WordNet, adjectives can be divided into 2 clusters i.e. one with same polarity and other one containing opposite polarity. Bootstrapping is another method used for Aspect (or feature) Related Terms (ART) identification. In [22], researchers' considers two types of ARTs for study-(1) nouns, adjectives, adverbs and verbs and (2) multiword terms. A new algorithm i.e. multi-aspect bootstrapping algorithm is used to segment multi-aspect sentence into single aspects. SenticWordNet 3.0 is a publicly available lexical resource used for supporting sentiment classification and opinion mining applications. The support vector machine (SVM) has also given good results for sentiment analysis.

#### B. *Statistical Approach*

N.Anwar, A. Rashid, S.Hassan [30] proposed a supervised information extraction system which extracts features and associated opinions. Frequency Bayesian classification technique was used to calculate probability distribution. PMI (Point wise Mutual Information)

algorithm uses mutual information as measure of the strength of semantic association between two words. PMI-IR uses Point wise Mutual Information (PMI) and Information Retrieval (IR) to measure the similarity of pairs of words or phrases. The semantic orientation of a given phrase is calculated by comparing its similarity to a positive reference word with its similarity to a negative reference word [31].

### C. Intelligent Feature Selection Approach

The IFS approach uses a feature relation network (FRN).FRN uses two important syntactic n-gram relations- (1) subsumption and (2) parallel [32]. These two relations occur between two n-gram features categories. IFS can be also combined with larger feature sets for enhanced Opinion-classification performance.

## V. PARAMETERS FOR PERFORMANCE EVALUATION

For Aspect-based opinion poll performance analysis of opinion extraction and Aspect extraction, recall and precision these two measures are used. For opinion poll in terms of polarity, algorithm performance is measured using accuracy.

### A. Recall

In information retrieval, recall is the fraction of the documents that are relevant to the query that are successfully retrieved. In Aspect or feature extraction algorithm, recall will be the fraction of the relevant opinions or aspects that are relevant to the query that is successfully retrieved. So it can be calculated using following formula-

Recall=

$$\frac{|{\text{relevant opinions /aspects}} \cap {\text{retrieved opinions /aspects}}|}{|{\text{retrieved opinions /aspects}}|}$$

### B. Precision

While retrieving any information, precision is the fraction of retrieved documents that are relevant to search. Similarly, using an opinion or aspect extraction algorithm precision is the fraction of retrieved opinions/ aspects that are relevant to search.

Precision

$$\frac{|{\text{relevant opinion s/aspects}} \cap {\text{retrieved opinions /aspects}}|}{|{\text{retrieved opinions /aspects}}|}$$

### C. F-measure

A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

$$F = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

### D. Accuracy

For sentiment analysis, the accuracy is the proportion of true results (both true positive opinions and true negative opinions) generated from the algorithm implemented on the review database. To clear the context by the semantics, it is often referred to as the "Rand Accuracy". It is a parameter used for the test.

An accuracy of 100% means that the measured values are exactly the same as the given values.

## VI. CONCLUSION

Opinion mining is in itself a vast research area Opinion mining reveals the concepts of text mining and also the concepts of information retrieval. In this survey paper it problems and Assigning weights to aspects. For Aspect-based opinion mining different tools like Rapid Miner, WordNet, SentiWordNet, etc are used. Natural Language processing, Computational Linguistics and text analytics can be applied to extract the subjective information from source review database and classify the polarity of the opinion stated. In recent years, automated content analysis is becoming important in opinion mining field. It is seen that opinion mining play an important role to help customers and business organizations to make decision about product /services. Mixed opinion problems is the area to be worked on in opinion mining

## ACKNOWLEDGEMENT

The author wish to express her sincere gratitude to Dr.G.K.Kharate, Principal, Dr.Varsha.Patil madam, Vice-Principal, and Prof.N.L.Bhale Sir, HOD, IT Department, Matoshri College of engineering, Nashik for their constant encouragement, important comments and helpful suggestions. The author is also thankful to all friends and family for their support. This paper is submitted with subject to University of Pune.

## REFERENCES

- [1] G. Carenini, R. Ng, and A. Pauls, Multi-Document Summarization of Evaluative Text, Proc. 11th Conf. European Chapter of the Assoc. for Computational Linguistics, pp. 305-312, 2006.
- [2] F.Y.Y. Choi, Advances in Domain Independent Linear Text Segmentation, Proc. First Meeting North Am. Chapter Assoc. For Computational Linguistics, pp. 26-33, 2000.
- [3] K.W. Church, Char Align: A Program for Aligning Parallel reviews Texts at the Character Level, Proc. 31st Ann. Meeting Assoc.For Computational Linguistics, pp. 1-8, 1993.
- [4] K. Crammer and Y. Singer, Pranking with Ranking, Proc. Neural Information Processing Systems, pp. 641-647, 2001.
- [5] M. Hu and B. Liu, "Mining Opinion Features in Customer Reviews," Proc. 19th Nat'l Conf. Artificial Intelligence, 2004.

- [6] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 168-177, 2004.
- [7] B. Liu, M. Hu, and J. Cheng, "Opinion Observer: Analyzing and Comparing Opinions on the Web," Proc. Int'l Conf. World Wide Web, pp. 342-351, 2005.
- [8] S.-M. Kim and E. Hovy, "Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text," Proc. Workshop Sentiment and Subjectivity in Text, 2006.
- [9] N. Kobayashi, K. Inui, and Y. Matsumoto, "Extracting Aspect-Evaluation and Aspect-of Relations in Opinion Mining," Proc. Joint Conf. Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 1065-1074, 2007.
- [10] M. Krauthammer and G. Nenadic, "Term Identification in the Biomedical Literature," J. Biomedical Informatics, vol. 37, no. 6, pp. 512-526, 2004.
- [11] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, vol. 2, nos. 1/2, pp. 1-135, 2008.
- [12] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," Proc. ACL-02 Conf. Empirical Methods in Natural Language Processing, 2002.
- [13] A.M. Popescu and O. Etzioni, "Extracting Product Features and Opinions from Reviews," Proc. Conf. Empirical Methods in Natural Language Processing, 2005.
- [14] G. Qiu, B. Liu, J. Bu, and C. Chen, "Expanding Domain Sentiment Lexicon through Double Propagation," Proc. Int'l Joint Conf. Artificial Intelligence, pp. 1199-120, 2009.
- [15] E. Riloff and R. Jones, "Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping," Proc. 16th Nat'l Conf. Artificial Intelligence 1999.
- [16] E. Riloff, S. Patwardhan, and J. Wiebe, "Feature Subsumption for Opinion Analysis," Proc. Conf. Empirical Methods in Natural Language Processing, pp. 440-448, 2006.
- [17] M. Thomas, B. pang, and L. Lee, "Get Out the Vote: Determining Support or Opposition from Congressional Floor-Debate Transcripts," Proc. Conf. Empirical Methods in Natural Language Processing, pp. 327-335, 2006.
- [18] I. Titov and R. McDonald, "A Joint Model of Text and Aspect Ratings for Sentiment Summarization," Proc. Assoc. for Computational Linguistics, pp. 308-316, 2008.
- [19] B. Wang and H. Wang, "Bootstrapping Both Product Properties and Opinion Words from Chinese Reviews with Cross-Training," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence, pp. 259-262, 2007.
- [20] D. Yarowsky, "Unsupervised Word Sense Disambiguation Rivaling Supervised Methods," Proc. 33rd Ann. Meeting Assoc. for Computational Linguistics, pp. 189-196, 1995.
- [21] J. Zhu, H. Wang, B.K. Tsou, and M. Zhu, "Multi-Aspect Opinion Polling from Textual Reviews," Proc. ACM Conf. Information and Knowledge Management, pp. 1799-1802, 2009.
- [22] Jingbo Zhu, H Wang, M Zhu, K. Tsou, "Aspect-based opinion polling from customer reviews", IEEE Transactions on Affective Computing, Vol. 2, no. 1, an-March 2011.
- [23] E. Riloff, J. Wiebe, and T. Wilson, "Learning Subjective Nouns Using Extraction Pattern Bootstrapping," Proc. Seventh Conf. Natural Language Learning at HLT-NAACL, 2003.
- [24] T. Zagibalov and J. Carroll, "Unsupervised Classification of Sentiment and Objectivity in Chinese Text," Proc. Third Int'l Joint Conf. Natural Language Processing, pp. 304-311, 2008.
- [25] B. Snyder and R. Barzilay, "Multiple Aspect Ranking Using the Good Grief Algorithm," Proc. Human Language Technology Conf. North Am. Chapter Assoc. of Computational Linguistics, pp. 300-307, 2007.
- [26] B. Pang and L. Lee, "Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales," Proc. Ann. Meeting on Assoc. for Computational Linguistics, pp. 115-124, 2005.
- [27] J.C. Reynar, "An Automatic Method of Finding Topic Boundaries," Proc. 32nd Ann. Meeting Assoc. for Computational Linguistics, pp. 331-333, 1994.
- [28] P. Fragkou, V. Petridis, and A. Kehagias, "A Dynamic Programming Algorithm for Liner Text Segmentation," J. Intelligent Information System, vol. 23, no. 2, pp. 179-197, 2004.
- [29] Xiaowen Ding, Bing Liu, Philip S. Yu, "A Holistic Lexicon-Based Approach to Opinion Mining", WSDM'08, 2008
- [30] N.Anwar, A. Rashid, S.Hassan,"Feature Based Opinion Mining of Online Free Format Customer Reviews Using Frequency Distribution and Bayesian Statistics", IEEE.
- [31] P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. "In Proc. of the Meeting of the Association for Computational Linguistics (ACL'02),2002.
- [32]. Ahmed Abbasi, P"Intelligent Feature Selection for opinion classification", IEEE 2010.