Feature-based Opinion Summarization: A Survey

Nirali Makadia

P.G Student of Computer Engineering

Marwadi Education Foundation Group of Institutions, Rajkot, India

_____

Abstract: **Significant advancement in e-commerce and Web 2.0 has led to the invention of several websites selling products online. These websites also facilitate the buyers to express their opinions about the products & their features in the form of reviews. Knowing these opinions and the related sentiments plays a vital role in decision making processes involving regular customers to executive managers. But these reviews are available in huge numbers hence referring them becomes a practically impossible task to achieve. Thus a new orientation of natural language processing called Opinion Mining & Summarization has emerged to deal with the problem. Feature-based Opinion Summarization is one of these summarization techniques which provide brief yet most relevant information about different features related to the target product. This paper represents an in-depth literature survey of various techniques for Feature-based Opinion Summarization.**

_____

I. INTRODUCTION

Online services play a very crucial role in every individual's day to day schedule. These services include daily news, weather forecast, banking transactions, shopping, social networking, blogging, and much more. With the rapid expansion in web technologies, online buying and selling of products has increased to a great extent. Added to the growth is the capability of users to share their feeling of satisfaction or criticism in the form of reviews. Knowing these opinions and its associated sentiments is important since it greatly affects the decision-making of an individual or an organization management system.

Looking at the current scenario, each product sold online nearly receives thousands of opinions from different users across the world. Hence going through this large number of reviews is a laborious task. On the other hand, referring only a few of them would lead to a biased decision. Thus opinion mining, sentiment analysis and summarization become a serious necessity. Opinion mining is simply finding sentences carrying an opinion for something whereas sentiment analysis is determining the positivity or the negativity associated with an opinion. Summarization is a way of presenting large amount of information using limited words still maintaining its meaning and relevancy. Similarly opinion summarization illustrates a summary for large number of opinionated sentences.

Feature-based Opinion Summarization is one of the opinion summarization techniques which provide brief yet most important information containing summary about different aspects related to the target product. Since it focuses on different features instead of giving the general details about a product, it has become more significant and demanded form of summarization. This technique is also known as Aspect-based Opinion Summarization. It is actually a way of generating summaries for a set of aspects or features of a specific product.

This paper is structured as follow: Section II describes the classification of opinion summarization techniques, section III explains the feature-based opinion summarization system, section IV contains literature review of different research papers related to Feature-based Opinion Summarization and at last section 5 holds the limitation of current system and future scope in the field of Feature-based Opinion Summarization.

II.   CLASSIFICATION OF OPINION SUMMARIZATION TECHNIQUES

The opinion summarization approaches can be broadly classified into aspect (feature)-based opinion summarization and non-aspect based opinion summarization. The figure 1 illustrates detailed classification of opinion summarization techniques.
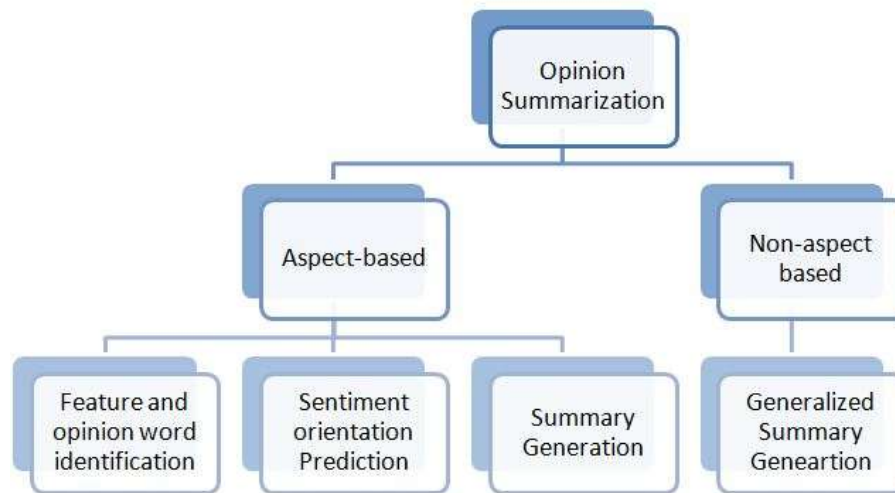
Figure 1: Classification of Opinion Summarization Techniques

Aspect-based Opinion Summarization technique concentrates on the "features" of the target product to shape-up the output summary. That is, it provides information about different features of a product instead of simply rendering a generalized scoop of information. This approach is being in great demand nowadays because it exactly shows what a customer usually tries to search while referring the reviews.

Non-aspect based Opinion Summarization technique produces a generalized summary over any target without considering its aspects or features. This summary is not bounded to any particular form or format. Different forms of these opinion summaries available include contrastive, abstractive, multi-lingual, entity-based, etc.

III. FEATURE-BASED SUMMARIZATION (FBS) SYSTEM

It is a technique that works around various features of a particular target and provides brief yet complete details about them. These features are the dominant factors of the text being processed. It comprises of three major explicit steps – feature/aspect and sentiment identification, sentiment orientation prediction, and summary generation. Figure 2 demonstrates an example of FBS system.

Product : Mobile Phone	
(Feature, Sentiment) Identification	(battery life, long) (speaker quality, great) (cost, expensive)
(Sentiment, Orientation) Prediction	(long, positive) (fine, positive) (expensive, negative)
Summary Generation (Star Ratings)	Battery Life : ***** Speaker : *** Cost : **

Figure 2: An Example of feature-based opinion summary generation

The feature identification process is finding the salient features of an entity, a topic or a product whereas sentiment identification means to recognize the text bearing sentiments associated with the identified features. The sentiment prediction step informs about the positivity or negativity of the sentiments identified on the features extracted in the first step. Finally summary generation step is executed to yield a potent summary by processing the results gained from the previous two steps. Table 1 enlists the techniques to perform the above mentioned processes.

Table 1: Techniques for different stages of Feature-based Opinion Summarization

Feature and Sentiment Identification	NLP Techniques
Feature and Sentiment Identification	NLP Techniques
	Mining Techniques
Sentiment Orientation Prediction	Lexicon-based Methods
	Learning-based Methods
Summary Generation	Statistical Summaries
	Text Summaries
	Aggregated Ratings
	Graphical Summaries

IV.   LITRETURE REVIEW

In [1], Bing Liu et al. described a model where the task of feature-based opinion summarization is performed by first mining the product features that have been commented on by customers using association mining technique, then identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative using a set of seed adjectives along with their orientations that grows later using WordNet and finally summarizing the results.

T. Ahmad et al. [2] developed an opinion mining system where the features and opinions are extracted using semantic and linguistic analysis of text documents; the polarity of the opinion sentences is discovered using polarity scores given by SentiWordNet and the generated summary is presented using a visualization module in a comprehendible way.

L. Zhao et al. in [3] introduced a fine-grain approach for opinion mining, which uses an ontology structure as an essential part of the feature extraction process by taking into account the relations between concepts. The approach involves data processing which includes POS-tagging and word segmentation. Then feature extraction is performed with integrated ontology that boosts the process accuracy. Then polarity identification is performed with the help of SentiWordNet and finally sentiment analysis is done to deal with negation and other semantics.

In [4], W. Zhang et al. developed a system called Weakness Finder that helps the manufacturers to find their product weakness from Chinese reviews by using aspect based sentiment analysis. The system extracts and group explicit features by using Morpheme based method and Hownet based similarity measure. Next it identifies and groups implicit features with collocation selection method for each aspect. Finally the polarity is determined by sentence based sentiment analysis method. The authors here assumes that the weakness of any product can be found out easily because the weakness would be probably the most unsatisfied aspect discussed in the customers' reviews.

A. Dengel et al. in [5] presented an extractive aspect-based sentiment summarization system which consist of an aspect detector for feature extraction that occurs frequently, a clustering module to cluster all the documents that have the occurrence of same aspect word within them in one group, a hybrid polarity detection system along with their generated feature set for determining opinion orientation and a textual and graphical summary generator module which uses an unsupervised polarity detection and ranking algorithm developed by them for summary generation.

In [6], A. Bagheri et al. proposed a novel unsupervised and domain-independent model for detecting explicit and implicit aspects in reviews for sentiment analysis. In the model, first a generalized method is proposed to learn multi-word aspects and then a set of heuristic rules is employed to take into account the influence of an opinion word on the detected aspect. Second a new metric based on mutual information and aspect frequency is proposed to score aspects with a new bootstrapping iterative algorithm which works with an unsupervised seed set. Third, two pruning methods based on the relations between aspects in reviews are presented to remove incorrect aspects. Finally the model employs an approach which uses explicit aspects and opinion words to identify implicit aspects.

R. Kumar et al. in [7] provided a method to mine different product features and opinion words based on customer opinion expressed in the review using a semantic based approach based on typed dependency relations. They tried to identify frequently and infrequent features from the given customer review using the typed dependency relations between each word present in the sentence and the opinion lexicon consisting of a list of subjective positive and negative words. They also tried to resolve the problem of pronoun resolution by replacing pronoun with appropriate product feature. The authors considered adjectives, verbs and even nouns as opinion words during their identification. Finally a summary consisting of positive and negative opinion sentences related to each product feature is generated.

K. Bafna et al. in [8] proposed a dynamic, domain-dependent system for feature-based opinion summarization of customers' opinions for online products. Firstly, identification of features of a product from customers' opinion is done using association rule mining and probabilistic power equation. Next, for each feature, its corresponding opinions are extracted and their orientation (positive/negative) is detected after forming feature-opinion pair by assigning the opinion word to nearest feature. At last, feature-based summarizations of the reviews are generated.

In [9], M. Dalal et al. presented a semi-supervised approach for mining online user reviews to generate comparative feature-based statistical summaries. It includes phases like preprocessing, feature extraction, followed by sentiment classification and summarization. They performed basic cleaning tasks like sentence boundary detection and spell-error correction in the preprocessing phase. Then after performing POS tagging using Link Grammar Parser, frequently occurring nouns (N) and noun phrases (NP) are considered as the possible opinion features based on multiword approach which are extracted along with the associated adjectives describing them, as indictors of their opinion orientation. Once features and opinion words are extracted, the sentiment polarity of the opinions is determined using SentiWordNet.

In [10], D. Wang et al. developed SumView, a Web-based review summarization system, to automatically extract the most representative expressions and customer opinions in the reviews on various product features. The system focuses on delivering the majority of information contained in the review documents by selecting the most representative review sentences for each extracted product feature. Once the product reviews are crawled, POS tagging and stop word removal processes are performed. Then the term-sentence matrix is constructed where each row represents a term and each column represents a sentence. Product features are extracted using association rule mining which users can select as per their wish and requirement. Once selected, the proposed feature-based weighted non-negative matrix factorization algorithm is performed to group the sentences into feature relevant clusters. Finally, the sentence with the highest probability in each cluster is selected to be presented in the summary for each feature.

D. Toshniwal et al. in [11] proposed an Aspect Base Sentiment Analysis System (ASAS), which handles the context dependent opinion words. The authors used an online dictionary for classifying the context dependent opinion word and then used natural linguistic rules to assign polarity to them which later become a training data set. Next for classification of the remaining opinion

words, they used opinion words and features together because the same opinion word can have different polarity in the same domain for different features. Finally the system generated a short summary for a particular product based on each feature.

M. Zaveri et al. in [12] proposed a technique that extends the feature-based classification approach to incorporate the effect of various linguistic hedges by using fuzzy functions to emulate the effect of modifiers, concentrators, and dilators. The authors presented a Fuzzy Opinion Classification technique where the user reviews are classifies as *very positive, positive, neutral, negative, very negative.* For this classification, they first extract the features, associated descriptors, and hedges, then they identify the polarity and initial value of the feature descriptors based on SentiWordNet score and finally calculates overall sentiment score using fuzzy functions to incorporate the effect of linguistic hedges.

S. Joseph et al. in [13] proposed a new syntactic based approach for aspect level opinion mining which uses syntactic dependency, aggregate score of opinion words, SentiWordNet, and aspect table together for opinion mining of restaurant reviews. The core tasks involved are aspect identification, aspect based opinion word identification and its orientation detection. In the proposed method, aspect and the associated opinion words are extracted using dependency parsing, polarity of opinions is determined using SentiWordNet and finally adjective, adverb-adjective, adverb-verb combinations are produced which shows the positiveness/negativeness of each aspect.

In [14], Z. Hai et al. employed a corpus-statistics association measure to identify features, including explicit and implicit features, and opinion words from reviews. The authors first extract explicit features and opinion words via an association-based bootstrapping method (ABOOT) which starts with a small list of annotated feature seeds and then iteratively recognizes a large number of domain-specific features and opinion words by discovering corpus statistics association between each pair of words on a given review domain. Next they provided a natural extension to identify implicit features by employing the recognized known semantic correlations between features and opinion words.

The table 2 given below depicts various methods developed by different authors for feature identification, Sentiment Prediction and Summary Generation.

Table 2: Methodologies for different stages of Feature-based Opinion Summarization phases

	Feature Identification	*Sentiment Prediction*	*Summary Generation*
[1]	POS tagging and association mining	Lexicon-based: Wordnet	Statistical + Textual Summary
[2]	POS tagging, term frequency and inverse document frequency.	Lexicon-based: SentiWordNet	Graphical Summary
[3]	Morpheme based method and Hownet based similarity measure to identify explicit features. Collocation selection method to identify implicit features.	Sentence-based sentiment analysis algorithm.	Statistical Summary
[4]	Ontology dependent identification	Lexicon-based: SentiWordNet	Textual Summary
[5]	Manual aspect detection	Unsupervised Polarity Detection and Ranking algorithm + SentiStrength lexicon	Pie Chart
[6]	POS tagging, heuristic rules and bootstrapping iterative algorithm.	------------------	Statistical Summary
	Feature Identification	*Sentiment Prediction*	*Summary Generation*
[7]	Semantic based approach using typed dependency relations	Lexicon-based	Statistical + Textual Summary
[8]	Association mining and probabilistic model	Lexicon-based: SentiWordNet	Textual Summary
[9]	POS tagging, frequency multiword approach	Lexicon-based: SentiWordNet	Graphical Summary
[10]	POS tagging, term frequency and inverse sentence frequency (tf-isf)	Non-negative matrix factorization	Textual Summary
[11]	Association mining and probabilistic model	Online dictionary and linguistic rules	Textual summary
[12]	POS tagging, multiword approach with decomposition strategy	Lexicon-based: SentiWordNet	Graphical Summary
[13]	POS tagging, term frequency and inverse sentence frequency (tf-isf)	Lexicon-based: SentiWordNet	Statistical + Textual Summary
[14]	Association-based Bootstrapping method for explicit feature identification. Semantic correlations for implicit feature identification.	Lexicon-based: SentiWordNet	--------------------

V. CONCLUSION

With the growing trend of being online, opinions and reviews have become one of the prominent measures for making decisions. But as the volume of opinionated text is rapidly growing, it's mining and summarization has become a severe necessity. This paper has illustrated two important methods for opinion summarization, namely aspect-based and non-aspect based opinion summarization. Additionally, it also navigates through a detailed survey made on Feature-based Opinion Summarization Techniques.

Despite of a lot of research efforts, current FBS systems suffer from many limitations and have a scope for significant improvement. Major works done so far are prominently focused on identification of explicit reviews, opinion words, sentiment polarity, increasing the effectiveness of various parameters like accuracy, recall, precision, F-measure, etc. Still a more advanced research is required for detection of sarcasm, irony, slangs, emotions, implicit features, non-noun explicit features, non-adjective and non-verb opinion words, refining grammatical mistakes, spelling mistakes, context dependency, syntactic dependency, semantics, multilingual opinion mining and summarization.

VI. REFERENCES

[1] M. Hu and B. Liu, "Mining and summarizing customer reviews," *Proc. 2004 ACM SIGKDD Int. Conf. Knowl. Discov. data Min. KDD 04*, vol. 04, pp. 168-177, 2004.

[2] M. Abulaish, Jahiruddin, M. N. Doja, and T. Ahmad, "Feature and opinion mining for customer review summarization," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, pp. 219–224, 2009.

[3] L. Zhao and C. Li, "Ontology based opinion mining for movie reviews," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, pp. 204–214, 2009.

[4] W. Zhang, H. Xu, and W. Wan, "Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis," *Expert Syst. Appl.*, vol. 39, no. 11, pp. 10283–10291, 2012.

[5] S. A. Bahrainian and A. Dengel, "Sentiment Analysis and Summarization of Twitter Data," *Comput. Sci. Eng. (CSE), 2013 IEEE 16th Int. Conf.*, pp. 227–234, 2013.

[6] A. Bagheri, M. Saraee, and F. De Jong, "Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews," *Knowledge-Based Syst.*, vol. 52, August, pp. 201–213, 2013.

[7] R. K. V and K. Raghuveer, "Dependency driven semantic approach to Product Features Extraction and Summarization Using Customer Reviews," pp. 225–238, 2013.

[8] K. Bafna and D. Toshniwal, "Feature based Summarization of Customers' Reviews of Online Products," *Procedia Comput. Sci.*, vol. 22, pp. 142–151, 2013.

[9] M. K. Dalal and M. a. Zaveri, "Semisupervised Learning Based Opinion Summarization and Classification for Online Product Reviews," *Appl. Comput. Intell. Soft Comput.*, vol. 2013, pp. 1–8, 2013.

[10] D. Wang, S. Zhu, and T. Li, "SumView: A Web-based engine for summarizing product reviews and customer opinions," *Expert Syst. Appl.*, vol. 40, no. 1, pp. 27–33, 2013.

[11] H. Kansal and D. Toshniwal, "Aspect based Summarization of Context Dependent Opinion Words," *Procedia Comput. Sci.*, vol. 35, pp. 166–175, 2014.

[12] M. K. Dalal and M. a. Zaveri, "Opinion Mining from Online User Reviews Using Fuzzy Linguistic Hedges," *Appl. Comput. Intell. Soft Comput.*, vol. 2014, no. 1, pp. 1–9, 2014.

[13] T. Chinsha and S. Joseph, "A syntactic approach for aspect based opinion mining," *Semant. Comput. (ICSC), 2015 IEEE Int. Conf.*, pp. 24–31, 2015.

[14] Z. Hai, K. Chang, G. Cong, and C. C. Yang, "An Association-Based Unified Framework for Mining Features," *Acm TIST*, vol. 6, no. 2, 2015.

[15] K. Khan, B. Baharudin, A. Khan, and A. Ullah, "Mining opinion components from unstructured reviews: A review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 26, no. 3, pp. 258–275, 2014.

[16] H. Kim and K. Ganesan, "Comprehensive review of opinion summarization," *Illinois Environ. …*, pp. 1–30, 2011.