ASPECT-BASED OPINION SUMMARIZATION FOR DISPARATE FEATURES

Nirali Makadia[1], Arindam Chaudhuri[2], Saifee Vohra[3]

[1] PG Student, Computer Engineering Department, MEFGI, Gujarat, India
[2] Associate Professor, Computer Engineering Department, MEFGI, Gujarat, India
[3] Assistant Professor, Computer Engineering Department, MEFGI, Gujarat, India

ABSTRACT

Significant advancement in e-commerce 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 an important 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 called Opinion Mining & Summarization has emerged to deal with the problem. Aspect-based (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. Hence the approach is in great demand nowadays because it exactly shows what a customer usually tries to search while referring the reviews. This paper focuses on extraction of different kinds of features associated with a target entity. Current state of the art suggests that concrete techniques are highly required for identification of those features which are not clearly mentioned. Thus our prime target is to deliver a succinct solution for effective identification of implicit features along with the explicit ones based on the opinion words encountered in user reviews. This is achieved by first extracting and processing the explicit features and then using them for the identification of implicit features. Finally summarization of sentences containing both kinds of aspects is done.

Keyword: *Features, Implicit, Explicit, Opinions, Summarization*

1. 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. 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. It can be performed at various levels of granularity like at document level, sentence level or at aspect level. For document level mining, a document is considered as a single entity to be observed. Similarly for sentence level mining, a single sentence and for aspect level mining, different aspects of an entity are taken into consideration. Initial studies on opinion mining and summarization has focused on classification of all the opinions as either positive or negative and determining the final polarity of the entire document. But the problem at this level occurred since different parts of a document (i.e. different reviews) may deal with different issues. As a solution, researchers tried sentence level mining but still it is

error prone because within a single sentence, multiple opinions with different polarities regarding different aspects of the target entity may exist which are necessary to be studied for true knowledge extraction and summary generation. Thus a feature-based approach to opinion mining has become a necessity where target entities and their expressed features are extracted from the text and then the expressed opinions are analyzed for every feature. This summary making procedure primarily involve works like features identification of the target, opinion words (sentences) related to the identified features determination, polarity detection of the obtained opinion words and finally providing a relevant feature-based summary regarding the target product. The final summary generated can play an instrumental role in influencing a buyer's or any managerial decision. Looking at the current scenario, we can observe that major works done so far has focused on identification and extraction of explicit features. But problem persists when the opinionated sentences that imply features remain undetected i.e. the sentences that contain opinions for a particular feature of target entity which is not clearly determined. This paper will identify disparate features of target entity so that a legitimately accurate opinion summary can be can be designed and presented to target audience.

2. RELATED WORK

2.1 Association-based Bootstrapping Method

Z. Hai et al. [14] 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.

2.2 Co-occurrence Association Rule Mining Approach

Z. Hai et al. [15] have proposed a two-phase co-occurrence association rule mining approach to identify the hidden features. In the proposed system, the first phase is rule generation where for each opinion word occurring in an explicit sentence, a significant set of association rules is created using co-occurrence matrix. Whereas the second phase clusters the rule consequents to make the generated rules more robust. Next whenever new opinion word is encountered, the matched list of robust rules are used and the one having the feature cluster with the highest frequency weight is fired and the corresponding implicit feature is identified.

2.3 Classification-based Approach

L. Zeng et al. [16] have proposed a classification based approach for implicit feature identification. The authors used word segmentation, POS tagging, dependency parsing for rule based method to extract explicit feature-opinion pairs. Then the pairs are clustered and the training documents for each cluster are constructed. Finally implicit features are identified through classification based feature selection.

3. PROPOSED METHOD

The proposed framework has five major modules. These modules are Input (User Review Sentences), Explicit Feature and Opinion Word extraction, Implicit Feature Identification, Summary Generation and Output (Aspect-based Summary).

The figure 3.1 below shows a diagrammatic view of the proposed framework along with its modules and their flow of interactions.

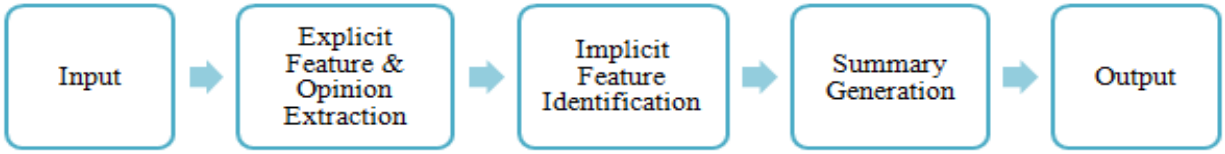


Fig. 3.1 Proposed Framework

For implicit feature identification, the process involve steps like sentiment orientation prediction, feature-opinion pair generation, replacing the synonym words with their corresponding feature word, counting the frequency occurrences of every unique pair and finally the identification of implicit feature.

Figure 3.2 below describes the steps performed for the identification of hidden features in an aspect-devoid review statement.

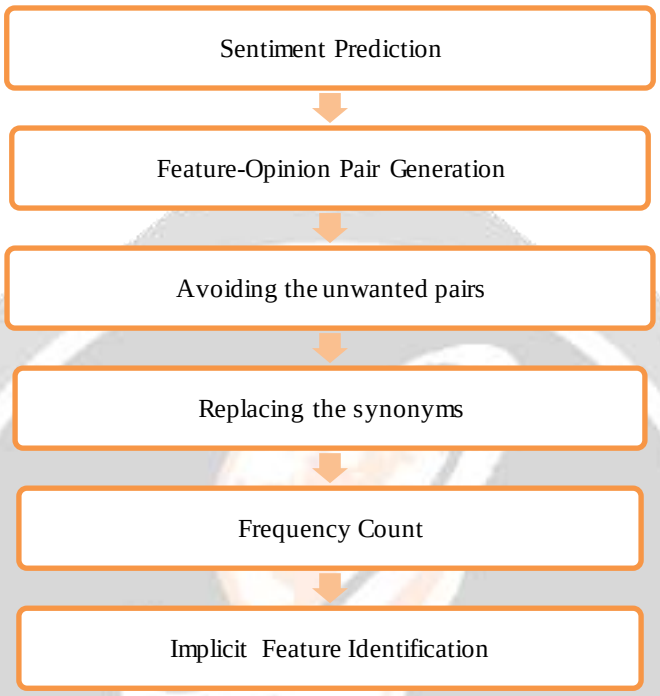


Fig. 3.2 Implicit Feature Identification Process

3.1 Sentiment Prediction

This step predicts the sentiment associated with a sentence. That is, it tries to identify whether the given sentence is positive or negative with respect to a product considered. It defines the score of 1.0 for positive statement and -1.0 for the negative statement.

3.2 Feature-Opinion Pair Generation

In this step, first of all the given sentence undergoes POS tagging where each word is tagged with its respective part of speech. Next the nouns and adjectives are filtered and stored in the form of feature-opinion pair.

3.3 Avoiding the unwanted pairs

While generation of these pairs, there are certain nouns which do not denote the feature words and hence are to be ignored. This issue is also considered where only the pairs containing feature words or the related synonyms are taken and the rest are ignored.

3.4 Replacing the synonyms

In this step, different synonyms for an aspect are replaced by their corresponding feature word. It is required to have uniformity and to avoid feature clustering.

3.5 Frequency Count

This step counts the occurrence frequency of each unique pair available. The uniqueness is defined by an opinion word. Hence based on an opinion word, its occurrence frequency for every feature, if pair is available, is calculated. This work is accomplished using RapidMiner, a tool available to perform certain data mining tasks.

3.6 Identification of Implicit Feature

In this step, an implicit feature is identified by comparing the frequency occurrence of the acquired opinion word with different features and selecting the one with highest count. If same frequency counts are obtained during comparison, then total frequency count of features is considered as the second check.

3.7 Summary Generation

Once the implicit features are identified and placed to their respective feature total count, an Aspect-based Summary will be generated where total number of positive and negative reviews will be displayed for every feature taken into consideration. Initially the total count will include only total number of explicit review sentences but the final total count will include total number of implicit and explicit sentences.

4. EXPERIMENTAL RESULTS

The results of the proposed system are compared to the results reported in IFLSABOOT [14], IF-LRTBOOT [14], CoAR [15], and CBA [16] with respect to the evaluation parameters considered. Comparison of all evaluation parameters used to analyze the proposed system performance relative to the existing methods is given in the table below. It is evident that the developed system provides better results.

Table 4.1 Comparative Analysis of Evaluation Parameters

Method	Precision	Recall	F-Measure
IF-LSABOOT [14]	72.99	69.93	71.43
IF-LRTBOOT [14]	66.06	63.29	64.64
CoAR [15]	70.75	62.59	66.42
CBA [16]	82.07	68.48	74.66
Proposed System	82.40	70.21	75.82

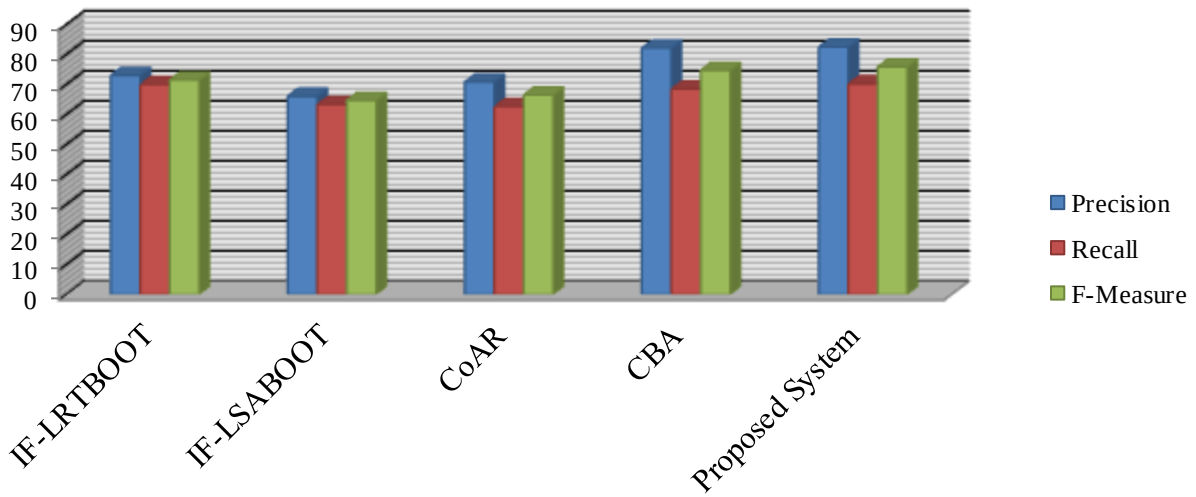
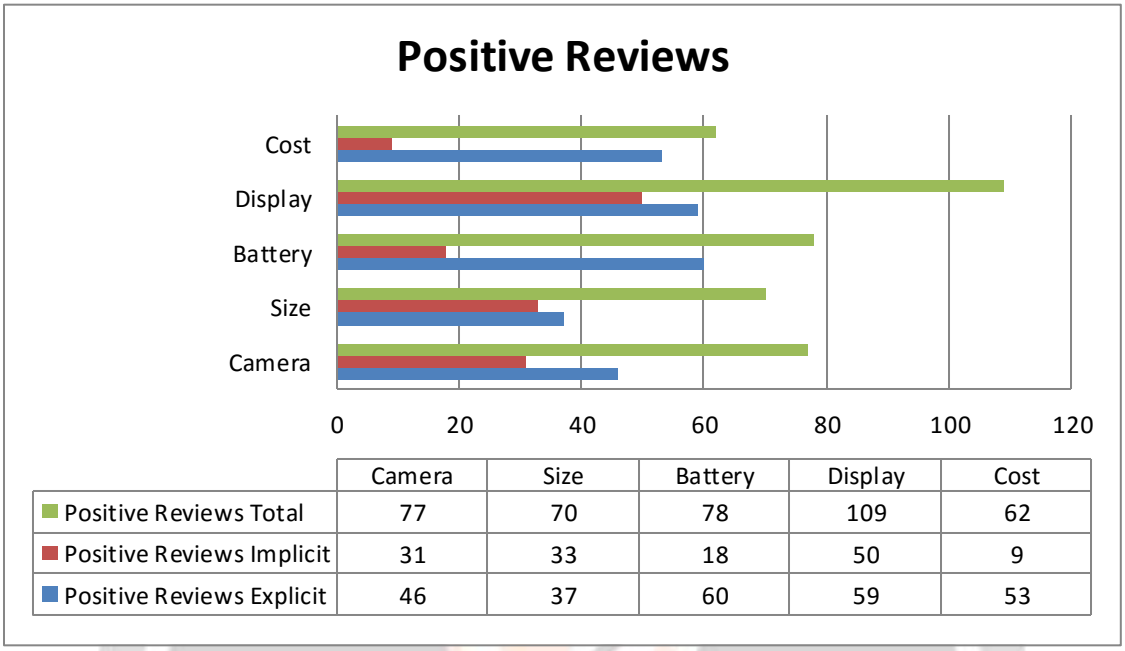


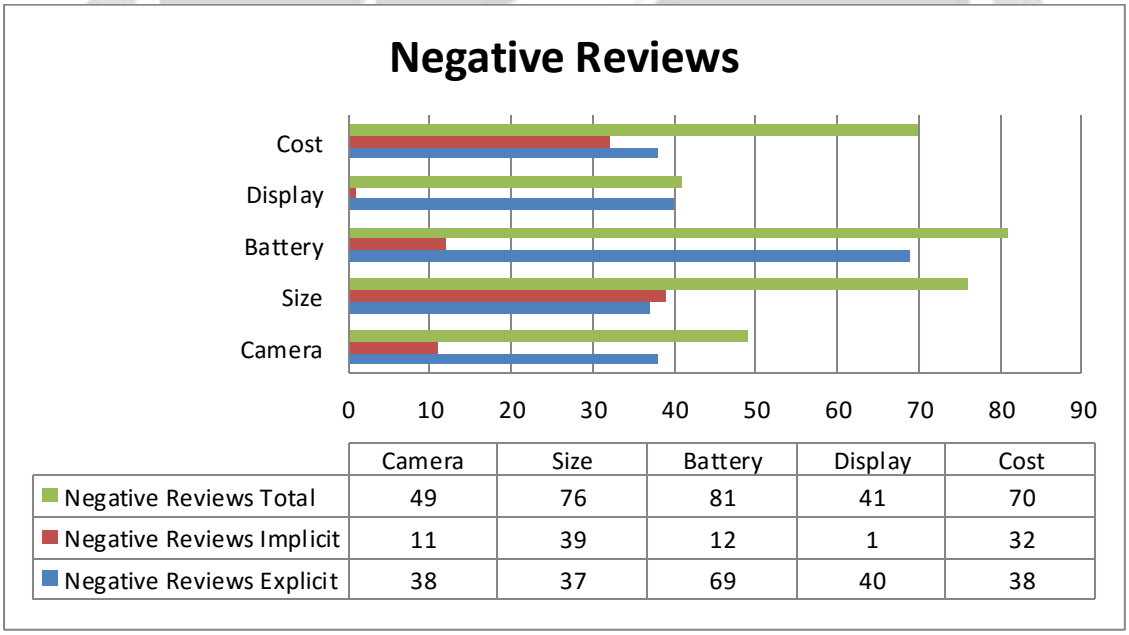
Fig 4.1 Comparative Analysis of Proposed System

The results obtained for explicit and implicit feature containing reviews involving both kinds of sentences, positive and negative, are illustrated in the figure given below. It is easy to note how the implicit feature identification and its addition to the process of summary generation makes the summary more productive.



	Camera	Size	Battery	Display	Cost
Positive Reviews Total	77	70	78	109	62
Positive Reviews Implicit	31	33	18	50	9
Positive Reviews Explicit	46	37	60	59	53

Fig 4.2 Results obtained for Positive Reviews



	Camera	Size	Battery	Display	Cost
Negative Reviews Total	49	76	81	41	70
Negative Reviews Implicit	11	39	12	1	32
Negative Reviews Explicit	38	37	69	40	38

Fig 4.3 Results obtained for Negative Reviews

The results obtained after subjecting the proposed method to two different datasets corresponding to two kinds of smartphones are depicted in the following table. These results are evaluated based on the standard parameters used through the entire process of analysis.

Table 4.2 Comparative Analysis of Proposed System for different datasets

Method	Precision	Recall	F-Measure
Smartphone 1	79.60	67.40	71.86
Smartphone 2	85.40	73.02	79.77
Average	82.40	70.21	75.82

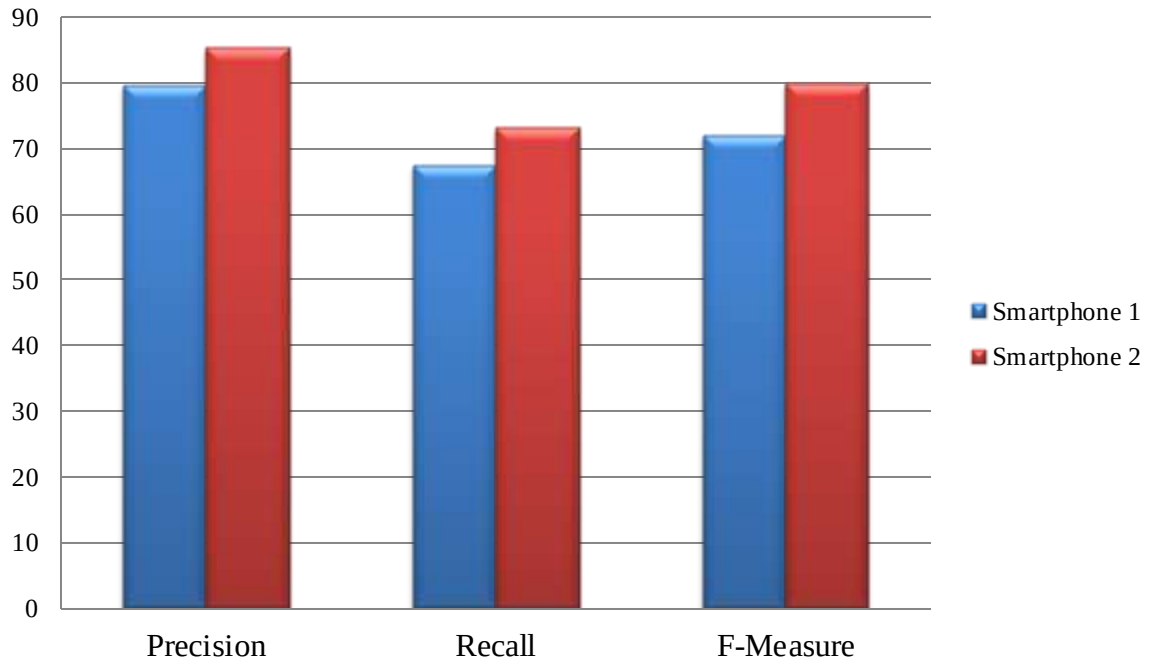


Fig 4.4 Results obtained for different datasets

5. CONCLUSIONS

Aspect-based Opinion Summarization is one of the recent yet very useful techniques for opinion summarization. This method tries to display the opinions or user reviews related to a product according to its various features (aspects). The current literature shows that existing systems works well with explicit kind of sentences but suffers a great deal of problems for identification and inclusion of implicit statements. The proposed framework identifies implicit features for a given opinion word and summarizes both kinds of sentences effectively. The current system illustrates a statistical summary of the user reviews. A summary by combining statistics with text can be generated making it more productive. The proposed framework can be made suitable for other domains as well by implying some modifications. Finally an enhancement to the system that covers verbs and nouns as well can be made to improve the overall performance of the system.

6. 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, p. 168, 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)*, vol. 5909, 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)*, vol. 5914, 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. –. 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, no. August, pp. 201–213, 2013.

[7] R. K. V and K. Raghuveer, "to Product Features Extraction and Summarization Using Customer Reviews," *Adv. Comput. Inf. Technol. Springer Berlin Heidelb.*, vol. 178, 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*, vol. 6, no. 2, 2015.

[15] Z. Hai, K. Chang, and J. Kim, "Implicit Feature Identification via Co-occurrence Association Rule Mining," *12th Int. Conf. CICLing , Tokyo, Japan.*, vol. 6608, pp. 393–404, 2011.

[16] L. Zeng and F. Li, "A Classification-Based Approach for Implicit Feature Identification," *12th China Natl. Conf. CCL 2013 First Int. Symp. NLP-NABD 2013, Suzhou, China, Oct. 10-12, 2013. Proc.*, vol. 8202, pp. 190–202, 2013.

[17] K. Schouten and F. Frasincar, "Finding Implicit Features in Consumer Reviews for Sentiment Analysis," *14th Int. Conf. ICWE, Toulouse, Proc. Pages*, vol. 8541, pp. 130–144, 2014.

[18] M. De Marneffe and C. D. Manning, "Stanford typed dependencies manual," no. April, pp. 1–28, 2015.

[19] 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.

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