



An Unsupervised Technique to Generate Summaries from Opinionated Review Documents

Ashwini Rao^(✉) and Ketan Shah

MPSTME, NMIMS University, Mumbai 400056, India
ashwini.rao@nmims.edu

Abstract. In the last few years, there has been a tremendous change in the way users behave over the net. This is mainly because of the growth that has happened in the field of Web technology. In earlier times, the role a user over the net played was that of an information consumer, now it's more of a data creator role. This role change has benefitted the world of politics, social network analysis, financial market analysis, etc., to name a few. Due to this huge creation of data, a mechanism that can automatically analyze and interpret this opinionated data is badly needed. Toward this research direction, unlike other summarization techniques, the paper proposes a novel method that is unsupervised and also domain-independent for generating opinion summaries. The final summaries that were generated are at four levels that range from being coarse to more granular ones. The proposed technique was tested on various data sets that were from nine different domains. The experimental results clearly indicated that 70–75% of the summaries generated were matching with the manually selected ones.

Keywords: Opinion summarization · Dependency parsers · Supervised and unsupervised learning

1 Introduction

In the past few years, Natural Language Processing (NLP) along with Internet technologies has seen a tremendous growth. This has managed to change the way the users now look and perceive the information on the net. In this era of immense social networking, the opinion of users over the net is now playing a central role in our decision-making process. A lot of research is being carried out to find out innovative ways of managing and analyzing this content generated by the user.

A lot of changes are also observed in the behavior of people in these social networking sites. They are now open to collaborate, share their views and opinions without any hesitation. Various fields such as finance, healthcare, online shopping, etc., are making use of this vast intelligence (in the form of opinions) collected over the net to make good business decisions. But the task is difficult to handle, due to the noisy and unstructured characteristics of the online data. This is the reason that many researchers are working

in this direction to develop certain automated techniques that can mine and analyze the content generated by the user. The subdomain of sentiment analysis which tries to give a solution to the above problem is opinion summarization. In this research domain, various methods and techniques are being researched to help users to collect, analyze, and draw conclusions from the huge collection of unstructured data. The growth in the various social networking sites such as Facebook, Ecommerce review websites, Twitter, etc., have also led to enormous wealth creation in terms of data. Analysis of this data can enable timely predictions and also help in planning better business strategies.

Toward this research direction, the paper proposes a novel technique for opinion summary generation. Unlike other approaches, the proposed technique is unsupervised as well as domain-independent and generates summaries at four different levels. The technique proposed was experimented on nine data sets that were from different domains.

The organization of the paper is as follows. Related research in the area of opinion summarization is discussed in Sect. 2. The proposed model of opinion summary generation is outlined in Sect. 3. Section 4 discusses the data sets used and experiments carried out. The future plan is outlined in Sect. 5.

2 Literature Survey

In recent years, many researchers have been trying to find probable optimized solutions to the different challenging problems that have emerged in the research domain of opinion summarization. One of these is the task of presenting the summaries to the user. It's a complicated task, as a summary that is informative to one set of users may not hold any value to that of the others. So, many researchers have been working on various techniques that display the summary of opinions in a way that the user can digest.

Many researchers have shown that it is always better to aggregate opinions from many sources or persons instead of just presenting the individual view of a person. Liu et al. [1] proposed one such type of summary which was structured and was called aspect-based summaries. These displayed positive and negative opinions about the set of features that frequently occurred in the reviews.

Cheng et al. [2] further worked on the same lines and came up with summaries that compared features of one product with that of the other product. These summaries proved to be useful to users who wanted to compare products across multiple feature dimensions. Other interesting types of summaries are the ones which try to show the perception of the user about a product over time. These trend tracking summaries helped the manufacturers to draw their team's attention toward products that were losing their popularity among users.

Many researchers [3–5] have even presented their summaries in a traditional fashion. By taking inputs that belong to an individual user or from a set of user reviews, textual summaries that are short and meaningful were generated. These summaries did not gain much popularity as any type of quantitative analysis could not be done and they were just suitable for the purpose of human reading.

Some researchers like Wei et al. [4] worked on approaches to determine the ways to select important features that would make their place in the final summary presented. The term frequencies of aspects/features were found out and further an appropriate threshold

was used to filter out some of them. They presented a summary at a brief and a detailed level. A brief level had just the headline featuring the sentence selected, and the detailed ones carried the sentiments of the previously selected sentences. Researchers [6] also presented summaries that highlighted the pros and cons of various features of a product. One more interesting summary was developed by Blair et al. [7]. In their approach, sentences were first rated based on the importance of features that they had. Then in the subsequent step, few of these highly rated sentences were selected for the final summary based on their length.

Some researchers [8] even used different methods based on graphs for generating opinion summaries. Several clusters based on aspects/features were first formed and within each of these clusters, few sentences were selected as leaders taking their measure of informativeness and usefulness into consideration. Few have even worked on generating summaries at different granular levels [9]. They start with a word-level summary, by ranking features as per their importance. Next is the phrase level of summaries that just displays short phrases that represent individual clusters. The final level is that of displaying highly ranked sentences where usefulness and informativeness measures are used for sentence ranking.

3 Proposed Approach to Generate Opinion Summaries

The paper proposes a novel way to generate opinion summaries at four distinct levels. In the research domain of opinion summarization, this phase of generating relevant opinion summaries is considered to be a challenging step by many researchers. It is mainly due to the fact that concise, precise, and visually effective summaries are easily understood and effectively used by the consumers for the process of decision-making. Many of the researchers, who are working in this direction, have used various graphical structures to display the summary results rather than a tabulated display. Also, many have used different sentence ranking methods to select sentences that would be part of the final summary. The challenge in these techniques is that they are domain-dependent as well as they require a good amount of training data before the model could be built.

Fig. 1 shows the approach for generating opinion summary. The process begins with the already extracted single and multiword features applying the automated rule-based algorithm [10, 11]. Then, by using the technique proposed [12], relevant feature opinion pairs were extracted. These features and their corresponding opinion pairs are used in the proposed domain-independent model to generate opinion summaries at various levels, which differentiates this method from the various other techniques for a summary generation.

As shown in Fig. 1, the initial step is to define a method to filter out few features that were extracted in the earlier phase of feature extraction [11]. It was found during experimentation that a larger number of relevant features were extracted. But, for the purpose of generating opinion summary, it would be better to filter out a few of them as our aim is to display only those review sentences that are important from the user's perspective. We go ahead with this, by the assumption that popular features are usually the ones that are mentioned frequently in the reviews. So, the first step is to find their frequency of occurrence and retain the top few features using an appropriate threshold.

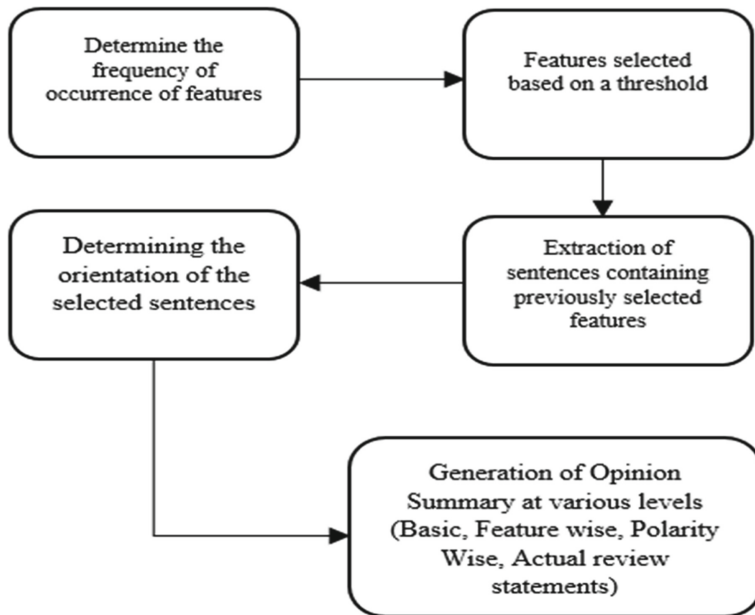


Fig. 1. Proposed technique for summary generation

The next step is to extract sentences having the selected features. The orientation of these sentences is then determined using a SentiWordNet dictionary [13]. This dictionary is used, as it has a huge collection of adjectives along with their orientation score. The orientation scores that are assigned to the words are based on the senses in which the words are used. As in the phase of feature extraction [10, 11], a dependency parser was used to parse the review documents, the context of adjective bearing words is easily available to determine the orientation of sentences.

So, now every sentence that is selected previously using the appropriate threshold is placed into a class having a positive, a negative, or a neutral polarity. With these polarities assigned to sentences, the summary is now generated from a coarse level to a more granular one which can be used as per user's need.

At the first level, the summary reveals details regarding the popular features mentioned in the reviews. The other information presented in this overall summary is the polarity of these features. Summary at the next level is termed as feature wise as it reveals more details about the set of features selected previously. This summary shows the opinion, i.e., positive or negative, of the author toward every individual feature. The third level is called average polarity wise summary. During experimentation, it was observed that the opinions for a few of the features that were selected were very strong even though their number remained small. To prevent losing even such granular detail which may generate better summaries, the average positive and negative polarity scores of every individual feature with its corresponding opinion was found out. This computed value was used in the display of selected features as per their average polarity score.

In the last level, the sentences bearing the features that were selected based on their average polarity score in the previous level was displayed. The above-discussed technique of generating opinion summary was experimented with nine other data sets that belong to the various domains like *Mobile phones*, *Automobiles*, *Hotel Industry*, and *Softwares*. The characteristic of the data set used along with the results of experimentation is discussed in the next section.

4 Experiments Conducted

The data sets that were used to evaluate the proposed technique were from various review sites such as Amazon, CNET, TripAdvisor.com, Carswale.com to name a few. Some of them were the golden data sets used by many researchers [1, 14] to showcase their contribution in the research field of sentiment analysis and few were manually crawled ones. Each of the reviews was around 8–10 sentences long, having an average of eight tokens per every individual sentence.

4.1 Results

As was discussed in the earlier section, the process of opinion summary generation using the proposed technique starts with the feature set and their corresponding opinion pairs that were extracted previously [10–12]. During experimentation, it was observed that the number of features in the category of relevant were too many. So, there was a need to limit their number. A literature survey conducted revealed that sentence ranking [2, 4, 6, 7] was one of the popular techniques used by many researchers to tackle the problem of dealing with the large feature vector. So, an initial step of filtering features was carried out to generate concise opinion summaries.

As discussed in the earlier section, the frequency of occurrence of features selected previously was found out. This was based on the assumption that popular features had more mentions in the reviews. For the automobile data set, there were around 45 features that got extracted in the feature extraction phase. Out of these, the ones that were having a good frequency of occurrence were the eight features that happened to be the top 20% of the sorted list. The review sentences corresponding to these selected features were then obtained. The selected top feature list of the automobile data set with frequency of occurrence is shown in Table 1. The table clearly indicates that features like *Seat Comfort*, *Leg space*, *Service Centre*, and *Sensor systems* are the most popular ones.

The next step as discussed earlier, was to extract sentences bearing these top-rated features and determine their polarity using the dictionary of SentiWordNet [13]. The orientation of few sample sentences having the features selected earlier is as shown in Fig. 2.

These sentences are then used in a subsequent step to generate summaries that start from a coarse level and extend up to a fine-grained level. Level 1 summary is at a higher level and gives details regarding the top-rated features and the number of opinions that are positive, negative, or neutral about the feature. Figure 3 shows the overall summary of the automobile data set.

Table 1. Features of automobile data set selected for summarization

Features	Frequency of occurrence
Seat Comfort	18
Leg space	15
Service Centre	14
Sensor Systems	11
Safety Feature	9
Brake System	8
Price Tag	5
Convenience Feature	4

The pricing is very high, even though it seems to give a tough competition-----pricing,high *****0.0, 0.425 POSITIVE ORIENTATION
They have service centres across the country and the service they provide is pretty good for the charges laid out-----service centres, good ***** 0.75, 0.2 POSITIVE ORIENTATION
The design of the engine looks shabby, would be better with more sophisticated safety features----- design, shabby *****0.5,0.85 NEGATIVE ORIENTATION
Seating at the back would have been much better but is still tolerable-----Seating, tolerable *****0.5,0.5 NEUTRAL ORIENTATION

Fig. 2. Sample sentences of automobile data set with orientation

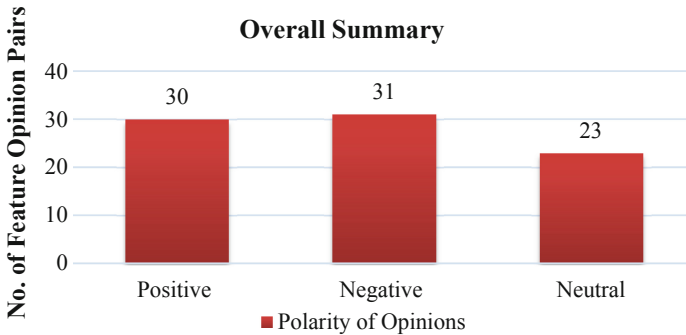


Fig. 3. Overall summary of the automobile data set

The figure clearly indicates that there are around 30 review sentences that bear positive opinion, 31 bearing a negative orientation, and 23 being neutral for a set of 8 top-rated features selected earlier.

The next level of summary for the automobile data set, termed as feature wise is shown in Fig. 4. The details regarding the number of positive, negative, and neutral opinions for each one of the features selected previously are shown here.

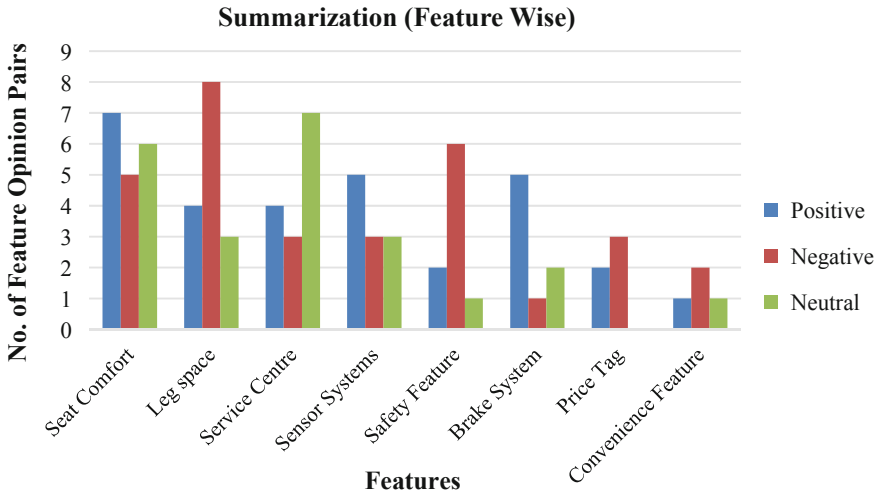


Fig. 4. Feature wise summary of the automobile data set

The figure is indicative of the fact that the features of *Leg space*, *Safety*, *Price tag*, and *Convenience* have a negative orientation, whereas that of *Service Centre* remains neutral and it is positive for *Seat comfort*, *Sensory*, and *Brake systems*.

The next level is termed as average polarity wise summary that presents much more detail than the previous two levels. Here the information regarding the average polarity scores that were computed for each of the feature opinion pairs that was previously selected is presented.

The importance of this granular level of the summary is in situations when for certain features even though the number of positive/negative opinions may be high, but their individual opinion strength may not be too great. For example, Fig. 4 indicates that feature of *Safety* and *Leg space* have more number of opinions that are with negative orientation. But the result displayed in Table 2 and Fig. 5 indicates that even though the number of negative opinions toward these features is high, they are still classified as positive. This is because of the fact that the opinion strength of the word bearing positive sentiment is more than that of the word with a negative sentiment.

Tackling such feature opinion pairs further increases the usefulness measure of the summary generated. The summary at the last level displays the actual review statements and is more granular. At this level, for the set of features selected, the average strength of its corresponding opinion that was computed in the previous level of the summary is used. The review sentences that are selected for the final display are the ones that have the score of polarity strength to be greater than the average score computed earlier. The level 4 summary of the Automobile data set is shown in Fig. 6.

Using the proposed technique, a similar set of summaries was also generated for the eight remaining data sets. The summary so obtained was not comparable with other similar approaches due to the following reasons that were evident during the conduction of literature survey.

Table 2. Average polarity wise summary of the automobile data set

Features	Positive polarity	Negative polarity
Seat Comfort	0.80	0.63
Leg space	0.79	0.71
Service Centre	0.75	0.00
Sensor Systems	0.51	0.52
Safety Feature	0.67	0.56
Brake System	0.63	0.00
Price Tag	0.58	0.31
Convenience Feature	0.88	0.00

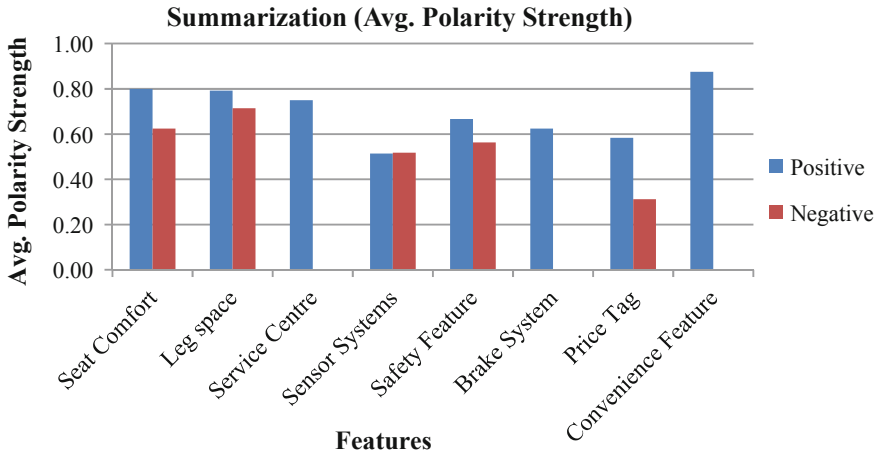


Fig. 5. Average polarity strength wise summary of the automobile data set

- Researchers evaluated the summaries that they generated based on the measure of informativeness, readability as well as non-redundant content of the summary.
- The researchers compared their summary generated with those that are stored in manually extracted files.
- Some researchers even based their evaluation according to the way in which polarity of every feature was distributed.
- Few of the researchers even argued that there can be no goodness measure of an opinion summary as it's very subjective. A summary that is found informative to one user may not be of any value to the others.

Seat Comfort	I own this car since Aug-2012 and one of the key differences is that I noticed though this is ONLY by reading and by viewing photographs is that the 3rd seating in this car seems to be better.
	My daily ride to office gets really irritating with the seat rattling in Ertiga.
	When I saw the vehicle today, except for the fact that I liked the ease of getting in and out of the third row, seating inside was horrible.
	In today's age what I can not digest is , Fixed head restraints , narrow and uncomfortable seats.
	The reclining and sliding seats mean that the vehicle would definitely be little more spacious and comfortable to use than its counterparts.
	Third row seats best for kids that too for short drives.
Leg space	The most glaring of all are those wafer thin seat cushions evident from the snaps.
	The model has a fixed middle row and the seating offers decent legroom for the third row. I was pretty dissatisfied with the car overall as it is not well packaged and I feel that vertical space is mismanaged, which could have created an altogether different experience otherwise.
Service Centre	Maruti Suzuki's sales & service network remains the unmatched numero uno with more choice , less arrogant dealers and open market availability of genuine spares.
	They have service centres across the country and the service they provide is pretty good for the charges laid out.
Sensor Systems	Even when encountered with apparent faults that cannot be blamed on anything else but poor design, or costcutting , or willful deficiency of service , buyers and users are willing to forgive and forget.
	In this the car the basic features like crash handling, airbags, parking sensors, speed sensitive and seat belts sensors, auto locking doors, crumple zones is just so bad, that its just not worth the money you pay for this big gain.
Safety Feature	One CD thing that baffles me is missing dog bar in Honda cars, I had an accident on the highway and condenser was dead meat.
	A car costing me close to 7.5 L on the road , and not having even basic stuff such as a driver airbag, is shocking.
	The only reason I dont want to go for this big car is because, somewhere in the back of my mind I had the feeling that safety features are virtually non existent in every car that Maruti makes in the sub lac bucket , hence whats the point in compromising safety.
Price Tag	Between the Ertiga and the Mobilio, I think the Ertiga looks much better from the front, and in all it has the great attributes of a safe, spacious family car.
	A well deserved 5 starts Pricing from Honda is disappointing compared to the product on offer.
	The pricing is very high, even though it seems to give a tough competition.
	Grow up Honda City, successful does n't mean people will buy anything at any price you give.

Fig. 6. Actual review statements summary of the automobile data set

5 Conclusion and Future Work

As the various techniques used for opinion summary generation are domain-dependent and supervised methods, the paper proposes a novel unsupervised domain-independent approach to this summary generation. The technique proposed generated summaries at four different levels revealing details about individual features from a basic level to that of a more granular one.

Level 1 summary termed as overall just displayed the total number of feature opinion pairs along with their polarity for individual feature selected. Level 2 summary had detail about every feature and number of positive and/or negative opinions that it carried. The level 3 summary had the features displayed based on the score of average polarity. The last level displayed the statements as per the filtering done in the previous steps.

The proposed technique was validated by running it through nine data sets that belonged to different domains. The summaries generated were tested for their readability and informativeness against files that had sentences that were manually selected for the final summary. It was observed that there was around 70%–75% matching of the summaries generated using the proposed method with that of manual ones.

Our future research direction is to come up with a way in which we can customize summaries as per the user needs. This can be made possible if we can think of a method to integrate the opinion summarization system with feedback from the user and use this information effectively in building an intelligent model.

References

1. Hu, M., Liu, B.: Mining opinion features in customer reviews. In: AAI 2004 (Vol. 4, No. 4, pp. 755–760)
2. Liu, B., Hu, M., Cheng, J.: Opinion observer: analyzing and comparing opinions on the web. In: Proceedings of the 14th International Conference on World Wide Web 2005 (pp. 342–351). ACM
3. Beineke, P., Hastie, T., Manning, C., Vaithyanathan, S.: Exploring sentiment summarization. In: Proceedings of the AAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications 2004 (Vol. 39). Palo Alto, CA: The AAI Press
4. Lun-Wei Ku, Y.-T.L., Chen, H.-H.: Opinion extraction, summarization and tracking in news and blog corpora. In: Proceedings of AAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs
5. Carenini, G., Cheung, J.C., Pauls, A.: Multi-document summarization of evaluative text. *Comput. Intell.* **29**(4), 545–576 (2013)
6. Zhuang, L., Jing, F., Zhu, X.Y.: Movie review mining and summarization. In: Proceedings of the 15th ACM International Conference on Information and Knowledge Management 2006 (pp. 43–50). ACM
7. Blair-Goldensohn, S., Hannan, K., McDonald, R., Neylon, T., Reis, G.A., Reynar, J.: Building a sentiment summarizer for local service reviews. In: Proceedings of World Wide Web (WWW-2008) Workshop on NLP in the Information Explosion Era 2008, (Vol. 14, pp. 339–348)
8. Seki, Y., Eguchi, K., Kando, N., Aono, M.: Opinion-focused summarization and its analysis at DUC 2006. In: Proceedings of the Document Understanding Conference (Duc) 2006 (pp. 122–130)
9. Ganesan, K., Zhai, C.: Opinion-based entity ranking. *Inf. Retrieval* **15**(2), 116–150 (2012)
10. Rao, A., Shah, K.: Model for improving relevant feature extraction for opinion summarization. In: Proceedings of IEEE International Advance Computing Conference (IACC) 2015 (pp. 1–5). IEEE
11. Rao, A., Shah, K.: An optimized rule based approach to extract relevant features for sentiment mining. In: Proceedings of 3rd International Conference on Computing for Sustainable Global Development (INDIACom) 2016 (pp. 2330–2336). IEEE
12. Ms. Rao, A., Dr. Shah, K.: A domain independent technique to generate feature opinion pairs for opinion mining. In: WSEAS Transactions on Information Science and Applications, 2018, (Vol 2, pp. 61–69)
13. Baccianella, S., Esuli, A., Sebastiani, F.: Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Lrec 2010 (Vol. 10, No. 2010, pp. 2200–2204)
14. Zhu, L., Gao, S., Pan, S.J., Li, H., Deng, D., Shahabi, C.: Graph-based informative-sentence selection for opinion summarization. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2013 (pp. 408–412). ACM