JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Using the Advanced Similarity Function to Create a Summary of Divergent Opinions

Pringal Gigaiya, Dhaval Nimavat
Research Scholar (M.Tech), Assistant Professor
Computer Engineering,
RK University, Rajkot, India

Abstract:

People utilise social media to research topics and make judgments based on what others have to say. A broad range of perspectives are shared through blogs, newsgroups, and other opinion-focused websites. As a result, consumers have a hard time relating to all of those points of view. The present opinion summary study focuses on distinguishing positive and negative viewpoints based on a set of criteria or characteristics. These viewpoints contain both positive and negative opinions on the product or issue at hand. The contrastive summary is built using basic level similarity metrics. This research tackles the problem by presenting a contrastive opinion assessment that shows both comments and views at the same time, allowing readers to make a comparison the two patterns of thinking. It enables correct viewpoint comparison within the context of a contrastive summary, therefore facilitates decision-making. The primary goal is to improve advanced sentence semantic similarity results while also summarising the counter-argument.

Keyword: Contradictory opinions, Contrastive summary, Advanced Similarity Function, Contrasting opinions

1. INTRODUCTION

Many parts of our daily lives are influenced by our opinions, including our behaviours, perceptions, and how we view and interpret the world. People may now easily express themselves on a range of topics by using platforms like opinion sharing websites. This is why, when faced with a difficult decision, we frequently seek the advice of others, thanks to the advancement of digital technologies. Users find it challenging to integrate all of the viewpoints accessible on a particular issue because there are so many. Opinion summarising, which generates a brief and consumable summary for a vast number of viewpoints, was used as a test.

When document-level mining is employed, a document is viewed as a single element to be monitored. Opinion summaries may be created at a variety of granularity levels, including phrase, label, and aspect. Initial studies on opinion mining and summarization focused on categorising all of the opinions as positive or negative and determining the resulting polarity of the entire content. Sentence-level mining considers a single sentence, whereas aspect-level mining considers several features of an entity. However, the user's issue is that in order to gain a clear understanding, he must browse through the list of favourable and negative opinions. People are really interested in reading about an issue since both aspects of it are reflected in a single frame. There are both favourable and negative opinions on the issue. As a result, the purpose of contrastive opinion summarising is to provide an auto-summary of a conflicting collection of terms on the same topic so that the user may quickly evaluate both sides of an argument.

A contrastive summary depicts both negative and positive perspectives on an issue, giving users a clear picture of the subject. Identifying such contrastive word pairings is a crucial initial step in creating contrastive opinion summaries. This contrastive opinionated description would be helpful for a user in comprehending both the positive and negative aspects of the given element. On several specific themes, traditional opinion summary methodologies culminated in the segregation of positive and negative viewpoints. The question now is what else can be done to have a better knowledge of the user after distinguishing positive and negative opinions. A descriptive summary is a method in which the user can see both angles of the coin at the same time and make a decision based on his or her preferences.

Imagine a person who posted reviews on his freshly obtained phone; he claims that the Smartphone ABC has a complex Operating System and so does not recommend purchasing it; however, the very same thing can be expressed affirmatively that the Smartphone ABC is extremely outstanding when it comes to technological applications. So, while both options are contrastive, they differ depending on the situation. This contrastive summary aids in the discovery of such viewpoints and allows the user to determine whether or not to purchase the phone. COS (Contrastive Opinion Summarization) help to identify opinions which looks most promising from different types of large amounts of opinions and summarise them using contrastive phrase pairs based on the common aspect in the COS assists in identifying the most significant opinions from a huge volume list of various types of opinions and generating a report reflecting contrastive phrase pairs, which is often the outcome of an opinion summarizer from the list of positive and negative reviews obtained.

In this research, we developed a system which would produce opposing opinion pairings, making the work of absorbing all of the user's perspectives for a specific topic/aspect efficient and less expensive consuming. Cosine similarity measurements are used in this study to determine the most comparable and contrastive pair. As far as we know, the movie reviews dataset, which includes both aspects and opinions, is not publicly available. We generated a new film reviews dataset for several components of the film, each with its own set of comments.

The remainder of the research is structured as follows: Section 2 covers current work on contrastive summary, Section 3 defines the problem, Section 4 describes the proposed method, Section 5 covers conducted experiments with results, and Section 6 describes about conclusion with future improvements.

2. LITERATURE REVIEW

In this section, we have reviewed earlier literature work in the area of contrastive opinion summary.

Opinion summarization entails categorising opinions related to emotional predictability and providing a contrastive summary that emphasizes positive and negative viewpoints. The topic of this assignment is opinion summaries in broad. In this arena, a great deal of research has indeed been done. The task and challenges of general opinion summarization are discussed in [13]. The study [14] provides a quick overview of opinion summarizing classification and also an overview of contrastive opinion summarization.

In [2], M. Ozsoy et al. presented research investigation, a unique approach called Contrastive Max-Sum Opinion Summarization (CMSOS) that aggregates the most contrastive and representative judgments was used. This approach provides a list of the most significant sentences for a specific element or topic. A more complex similarity function is the Cosine Similarity function, is employed in this model for calculating sentence similarity, together with Term Frequency (TF) and Inverse Term Frequency (ITF). The mixture of Similarity measures and frequency approaches yields a superior outcome.

H. D. Kim et al. in [3], presented a technique known as contrastive opinion summary in their study. The goal of COS would be to find the most similar sentences that include opposing viewpoints. Content similarities and contrastive similarity measurements are two generic methodologies based on similarity measures provided by the author. Contrastive similarity calculates content or opinions in two separate groups of reviews, whereas content similarity calculates content or opinions in the same group of reviews. Two algorithms are employed in this paper: the Contrastiveness-First (C-F) method and the Representativeness-First (R-F) algorithm. In terms of precision and aspect coverage, the C-F approach performs better. For better outcomes, enhanced semantic-based similarity measures might be applied.

R. Sipos et al. in [7], presented a strategy for selecting matches of excerpts from reviews in order to construct a summarising product comparison summary in a research study. They gave a suggestion for the submodular objective function, which is paired with the snippet. Because product opinions are readily available, extracts from them are included here.

By using customer feedback on the given pairs, they were able to understand simplification across diverse product pairs using a supervised learning approach. Like COS, the model generates a summary by comparing excerpts from two separate products.

J. Guo et al. in [1], improved the shortcomings of conventional contrastive opinion summarization (COS) by combining expert and ordinary opinions. For those subjects that cause debate, Expert Guided COS (ECOS) is an approach that employs Latent Semantic Analysis (LSA) to locate topics/labels from views, according to the author. For such contentious issues, the ECOS approach is used to select the most divergent argument pairings. This methodology generates a conclusion based on the factors and combines expert and ordinary opinions.

3. DEFINITION OF THE PROBLEM

When addressing themes where competing opinions have been expressed, recognising these opposing viewpoints, and analysing their progress through space and time, people are more sensitive. As a result, the summaries must include a variety of viewpoints. To allow direct comparison of different topics/products, a method for constructing concise and comparative summaries based on opinions is necessary. One of the answers to the problem is Contrastive Opinion Summarization. It presents a summary of differing viewpoints at the aspect level. To find contrastive pairs of concepts from positive and negative assessment is the main goal of this work, and then produce a contrastive summary for better user comprehension.

4. PROPOSED METHOD

There are five primary modules in the proposed framework. Sentiment prediction, Tagging of Film review, Film Reviews Input and selection of the most descriptive reviews from both positive and negative feedback among the most diverse series of reviews, Summary Generation and Output are the modules in question (Contrastive Opinion Summary).

Advanced similarity measures are used to identify from each set of positive and negative reviews, the most representative review and a collection of reviews that contrast.

The proposed framework, with its modules and their interactions, is depicted in Figure 1 as follows.

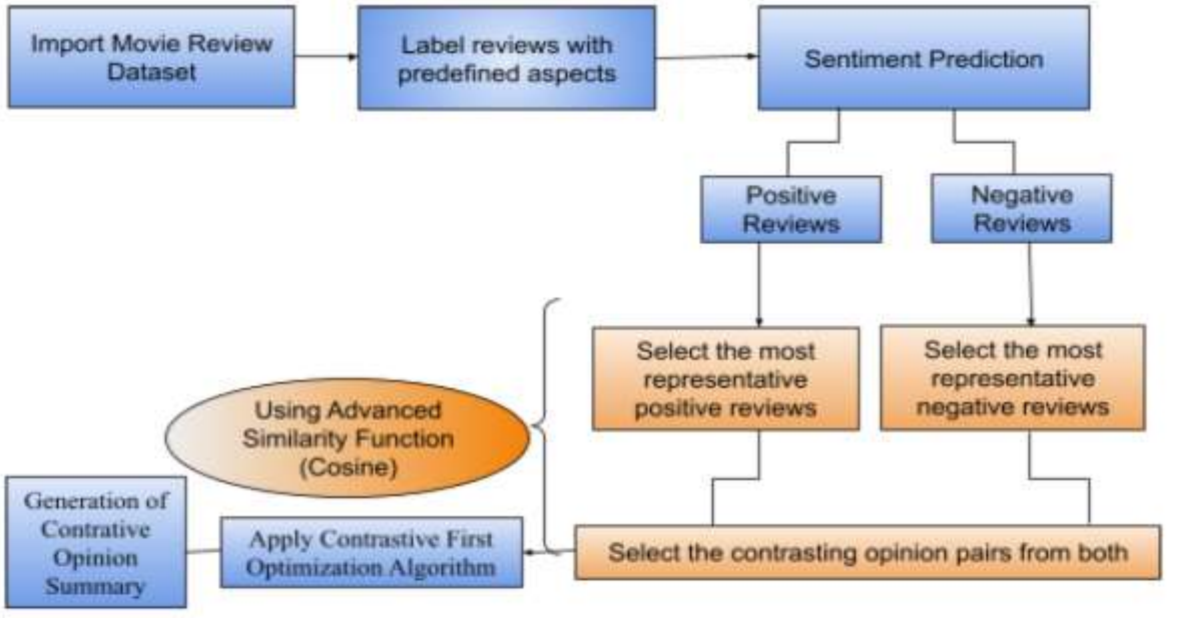

Figure 1: Proposed Framework

The Content Similarity Function is used to identify the most indicative review from both the positive and negative reviews, while the Contrasting Similarity Function is used to select the contrasting pair of reviews. Both similarity functions were calculated using Cosine Similarity, a more advanced similarity metric.

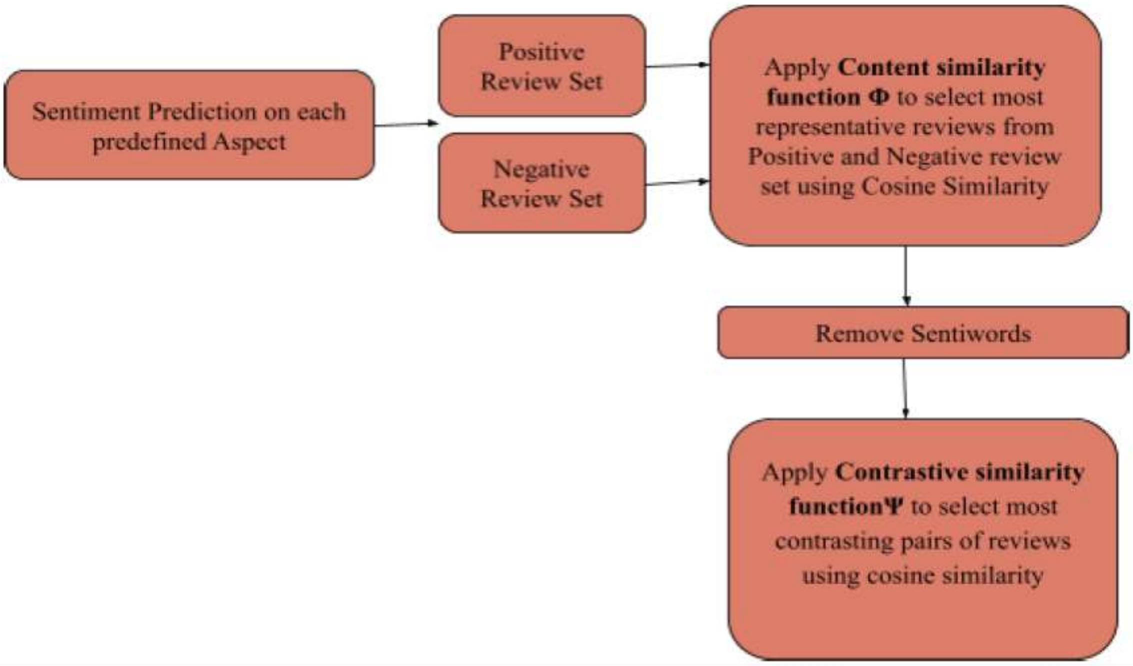

Figure 2: The Procedure for Calculating Similarity Measures

4.1 Predicting Public Opinion

The emotion connected with a sentence is predicted in this step. That is, it attempts to determine whether a given language is positive or negative in relation to the product in question. It assigns a 1.0 value to positive remarks and a -1.0 score to negative comments.

4.2 Most Representative Opinions

On a positive collection of concepts, the Content Similarity Function is determined using the advanced similarity metric, Cosine Similarity. Sentiment prediction elicits both good and negative feelings, and this approach is used to gather both.

The technique is repeated in the same way for the series of multiple opinions. The following is the formula for computing the Content Similarity Function [3]:

$$\emptyset(s_1, s_2) = \frac{\sum_{u \in s_1} max_{v' \in s_2} w(u, v') + \sum_{v \in s_2} max_{u' \in s_1} w(u', v)}{|s_1| + |s_2|}$$

…………………….. (1)

Where |s1| & |s2| are indeed the sum numbers of phrases s1 and s2, respectively, and w(u,v) has been the term similarity function. Content similarities and contrastive resemblance functions are computed using an approach based on cosine similarity [18] of feature axes. As per equation (2), a cosine similarity is used to determine the degree of the resemblance of two phrases.

$$\cos(a,b) = \frac{a.b}{\|a\|\|b\|} \quad \text{………………..…….. (2)}$$

The cosine value will be in the range [0,1], where 0 denotes the opposite direction and 1 denotes same direction for two strings and for two strings which are not equal. This cosine value is utilised to compute similarity functions in the future.

4.3 Choosing a Pair of Opposing Viewpoints

Choosing the most relevant reviews from each pair of positive and negative sentences yields the Contrastive Similarity Function. The contrastive combinations that will be included in the summary are returned by the Contrastive Similarity function. After emotive words like negations and adverbs are eliminated, the similarity is calculated using Cosine Similarity. The rest of the calculation is carried out in the same manner as for. The formula for the remaining of the computation is the same as for the Contrastive Similarity.

4.4 Pairs of opposing viewpoints are ranked.

After detecting the contrasting mixture of sentences, Contrastive First (CF) [3] is used to decide the sequencing of the contrasting combination of phrases in the final summary. This method is based on contractiveness as well as generalisability.

$$S^{*} = \arg\max \left(\lambda\, r(S) + (1-\lambda) c(S)\right)$$

$$S^{*} = \arg\max \left(\frac{\lambda}{|X|} \sum_{x \in X} max_{i \in [1,k]}\, \phi(x, ui) + \frac{\lambda}{|Y|} \sum_{y \in Y} max_{i \in [1,k]}\, \phi(y, vi) + \frac{1-\lambda}{k} \sum_{i=1}^{k} \psi(ui, vi) \right) \quad \text{…………………. (3)}$$

where (x,ui) is similarity among x means positive opinions, (y,vi) is similarity among y means negative opinions, (x,ui) is similarity among y means negative opinions and (ui, vi) is a contrastive pair of opinions. The opposing pair with the highest content similarity value is used for carrying out this task.

4.5 Generation Summary

Positive and negative ratings will be presented for each element studied in a Contrastive Opinionated Summary. The knowledge of best practises will be displayed in a tabular manner when the contrastive pair's order has been finalised, presenting differing perspectives on another element.

5. EXPERIMENTAL RESULTS

The dataset used during our experiment is discussed in this part, as well as the outcomes achieved.

5.1 DATASET DESCRIPTION

The real-time movie reviews data collection is utilised in our suggested system. These evaluations include both favourable and negative feedback from the audience inside the form of cheap or extensive comments. Some reputable internet movie review sites, such as rottentomatoes.com, were used to gather movie reviews. The real-time dataset is obtained using the import.io tool, and the manual tagging, with specific aspects, is done on reviews. Based on predetermined criteria, each representative movie reviews are divided into different categories. For the proposed method, the acquired reviews are divided into individual sentences.

Table 1: Dataset of movie reviews with sentiment prediction

ID	Movies	Postive Reviews	Negative Reviews	Total Reviews
1	Kick	25	19	44
2	Bajirao Mastani	28	27	55
3	Dilwale	29	28	57
4	Tenu Weds Menu	24	23	47
5	Fan	25	24	49
6	Kapoor and Sons	23	25	48
7	Prem Ratan Dhan Payo	22	24	46
8	Tamasha	20	25	45
Total				341

5.2 EVALUATION PARAMETERS

Precision, recall, and F-measure are the parameters that are used to evaluate the proposed system's performance. Precision is the first parameter, which would be defined as the fraction of relevant occurrences retrieved. The proportion of relevant instances collected is defined by another parameter called Recall. F-measure is used to check the proposed system's correctness. It is calculated from the average of both the recall and precision values. The mathematical model illustrates how to determine precision, recall, and F-measure:

$$Precision = \frac{|Relevant\ Sentence\ pairs\ \cap\ Retrieved\ Sentence\ pairs|}{|Retrieved\ Sentences\ pairs|} \quad \ldots\ldots\ldots\ldots\text{(4)}$$

$$\text{Recall} = \frac{|\text{Relevant Sentence pairs} \cap \text{Retreived Sentence pairs}}{|\text{Relevant Sentence pairs}|} \quad \ldots\ldots\ldots\ldots\text{(5)}$$

$$F - measure = \frac{2 \times Recall \times Precicion}{(Recall + Precicion)} \quad \ldots\ldots\ldots\ldots\ldots\text{(6)}$$

Relevant values are manually identified in order to evaluate these measures. The suggested system generates a summary of the opposing viewpoint pairs. The above-mentioned parameters are used to calculate movie reviews.

5.3 RESULTS

According to the results, the proposed method has of 77.6% of accuracy. The F-measure value, Recall value and Precision value are shown in following graph as per Fig. 3.

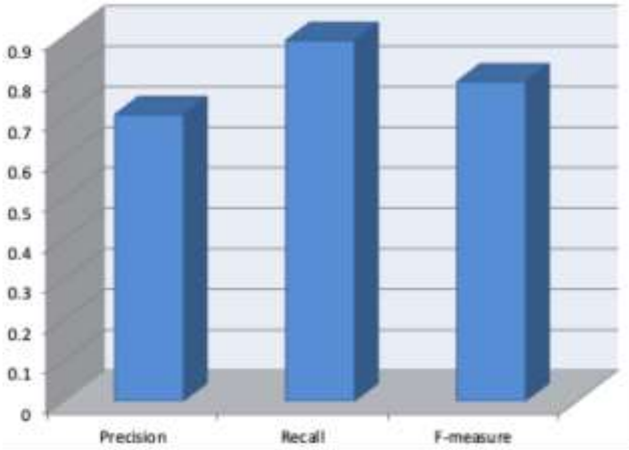

Figure 3: Analysis of the proposed system's results

In Table 2, we compare the outcomes of our technique to those given in [1] and [2].

Table 2: Comparison of results

Methods	Recall	Precision
Kim et al-RF	0.737	0.503
Kim et al-CF	0.804	0.537
CMSOS	0.889	0.649
Proposed System	**0.891**	**0.708**

The methods described in [1] are labelled Kim et al-RF, in which the process of clustering must be done before selecting the contrasting pairings which needed the use of an optimal clustering method. Kim et al-CF, in which basic similarity measures were used but did not result in a constructive contrastive summary.

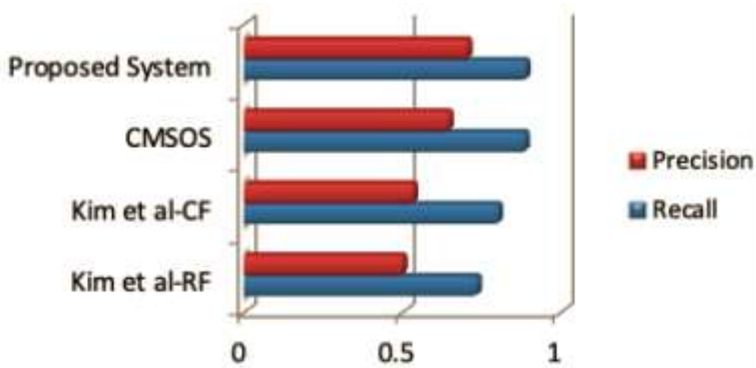

Figure 4: Analyse Existing Solutions Comparatively

In the CMSOS technique, either representativeness and contractiveness are used simultaneously. When employing cosine similarity metrics, the suggested method performs well because it offers the maximum similarity score regardless of text size. In the selection of contrasting pairs, the advanced similarity measure outperforms previous methods and offers a relevant summary.

6. CONCLUSION AND FUTURE WORK

The proposed work focuses on employing a collection of characteristics or qualities to distinguish positive and negative viewpoints. This study tackles the problem by providing a contrastive opinion evaluation that simultaneously shows both positives and negatives, allowing readers to compare the two ways of thinking. The major goal is to improve advanced sentence semantic similarity scores while summarising the opposing perspective at the same time. The new technique will automatically generate an educational contrastive summary of a large number of movie reviews and their elements. According to experiment results, the new approach uses the cosine similarity formula to produce superior results and can be utilised to build a contrasted summary. This proposed approach can be used for big real-time datasets in the future. One can use other Advanced Semantic-based similarity functions to compare the results.

7. REFERENCES

[1] J. Guo et.al. “Expert-Guided Contrastive Opinion Summarization for Controversial Issues,” Proceedings of the 24th ACM international conference on World Wide Web Companion, pp. 1105–1110, 2015.

[2] M. Ozsoy and R. Cakici,“Contrastive Max-Sum Opinion Summarization,” Information Retrieval Technology. Springer,InternationalPublishing,vol.1,pp.256–267, 2014.

[3] H.Kim and C.Zhai, “Generating Comparative Summaries of Contradictory Opinions in Text,” Proceedings of the 18th ACM conference on Information and knowledge management ,pp. 385–393, 2009.

[4] M.Camper and K.Jezek, “Comparing Semantic Models for Evaluating Automatic Document Summarization,” Text,Speech,andDialogue,Springer,InternationalPublishing,pp.252–260, 2015.

[5] M. Camper and K. Jezek, “Topic Models for Comparative Summarization,”Text, Speech, and Dialogue,Springer Berlin Heidelberg, pp. 568–574, 2013.

[6] M. J. Paul and R. Girju, “Summarizing Contrastive Viewpoints in Opinionated Text,” pp. 66–76, 2010.

[7] S. Ruben, and T. Joachims,"Generating comparative summaries from reviews," Proceedings of the 22nd ACM international conference on information and knowledge management , pp. 1853-1856, 2013.

[8] M. De Marneffe, “Finding contradictions in text,” Association for Computational Linguistics, vol.8, pp. 1– 25, 2005.

[9] S. Harabagiu, A. Hickl, and F. Lacatusu, “Negation , Contrast and Contradiction in Text Processing,” American Association for Artificial Intelligence,vol.6, pp. 755–762.

[10] P. Achananuparp, X. Hu, and X. Shen, “The Evaluation of Sentence Similarity Measures,” Proceedings of the 10th international conference on Data Warehousing and Knowledge Discovery, Springer Berlin Heidelberg, pp. 305-316, 2008.

[11] A. Popescu and M. Pennacchiotti, “Detecting Controversial Events from Twitter”

[12] H.D. 2010. ACM international conference on Information and knowledge management , pp.1873-1876,

[13] 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.

[14] H. Kim and K. Ganesan, “Comprehensive review of opinion summarization,” Illinois Environ., pp. 1–30, 2011.

[15] S. Bhattacharjee, A. Das, U. Bhattacharya, S. K. Parui, and S. Roy, “Sentiment analysis using cosine similarity measure,” Recent Trends Inf. Syst. (ReTIS), IEEE 2nd Int. Conf., pp. 27–32, 2015.

[16] A. Jeyapriya and C. S. K. Selvi, “Extracting aspects and mining opinions in product reviews using supervised learning algorithm,” 2nd Int. Conf. Electron. Commun. Syst. ICECS, pp. 548–552, 2015.

[17] J. Jin, P. Ji, and R. Gu, “Engineering Applications of Artificial Intelligence Identifying comparative customer requirements from product online reviews for competitor analysis,” Eng. Appl. Artif. Intell., vol. 49, pp. 61–73, 2016.

[18] S. Zhu, “TOP-K Cosine Similarity Interesting Pairs Search,” Fuzzy Systems and Knowledge discovery (FSKD),7thInternationalConference,vol.3,pp.1479–1483, 2010.

[19]M.Hu and B.Liu, “Mining and Summarizing Customer Reviews,” 2662 www.ijariie.com 3769 In Proceedings of the 10th ACM SIGKDD International Conference on Knowledge discovery and data mining, pp. 168-177, 2004.