

Evaluation Study on Gaussian Probability Distribution Based Feature Extraction for Opinion Mining from Customer Reviews

^IDr R.S. Vetrivel, ^{II}M. Arun Kumar, ^{III}P Jeyanthi Rani

^IProfessor, Computer Science, Subramanya College of Arts & Science, Palani.

^{II}Research Scholar, Computer Science, Subramanya College of Arts & Science, Palani

^{III}Assistant Professor, Computer Science, Subramanya College of Arts & Science, Palani

Abstract

Opinion mining is the analysis of features such as products, services and attributes in the form of document, sentence and word level that are extracted from various customers. The customer reviews are typically expressed in different formats namely sentence-level which classifies the features in an individual review documents. The existing opinion mining approach categorizes the opinion features only from online customer reviews across different corpus into dependent features and independent features. Intrinsic and Extrinsic Domain Relevance (IEDR) approach identifies only candidate features that are located in customer reviews and does not provide commonly presented features of the document. However, non-noun features are not able to identify from opinion features and it is difficult on feature extraction. Therefore, Gaussian Probability Distribution (GPD) is presented in proposed work to identify the Feature Extraction for Opinion Mining from Customer Reviews. On the developing world, online shopping offers the chance for customers to mark reviews about products. According to a Gaussian Probability Distribution, the comments in the form of text, speech and database source from different customer are distributed with each class. Natural Language Processing is used to summarize the customer review about the products and extract the duplicate reviews stored in the database. With the help of Naive Bayes classification and summarization technique, reviewer commands from the customers are classified as positive and negative opinions. The advantages of Gaussian Probability Distribution are extracting clear features at the sentence level of customer review and removing the unwanted features for opinion mining. Finally, the performance evaluation of proposed work is preformed in the metrics of precision rate, accuracy and feature extractions based on the number of review commands.

Key Terms

Gaussian Probability Distribution (GPD), Natural Language Processing (NLP), Naive Bayes classification and summarization technique (NBCST)

I. Introduction

1. Opinion Mining

E-commerce becomes more popular where numerous products are sold on the Web. People who are interested with any products are able to purchase them by using the Web. The outcome of customer satisfaction is more important for online merchants. Hence, online merchants facilitate the customers for reviewing or expressing their opinions and shopping experiences on the products that they purchase. Internet usage by people grows rapidly and higher number of users gives their reviews which help for other users to make a good decision. By using these received reviews on a product, a new user feels comfortable to know about that particular product. Popular products are able to obtain hundreds of reviews at some large merchant sites.

Product reviews given by experienced customers are difficult for the new customers to read those reviews through which making a decision on whether to buy the product. Therefore, opinion summary is obtained as the result of sentiment prediction which helps for the new customers. The process of sentiment prediction or classification produces opinion summaries from the customer reviews that provide a well-organized breakdown by aspects/topics, various formats of textual summaries and temporal visualization. There are different forms of summaries which are complement to one another by providing a different level of understanding. For example, sentiment prediction on reviews of a product results in a very general view of what the users feel about the product.

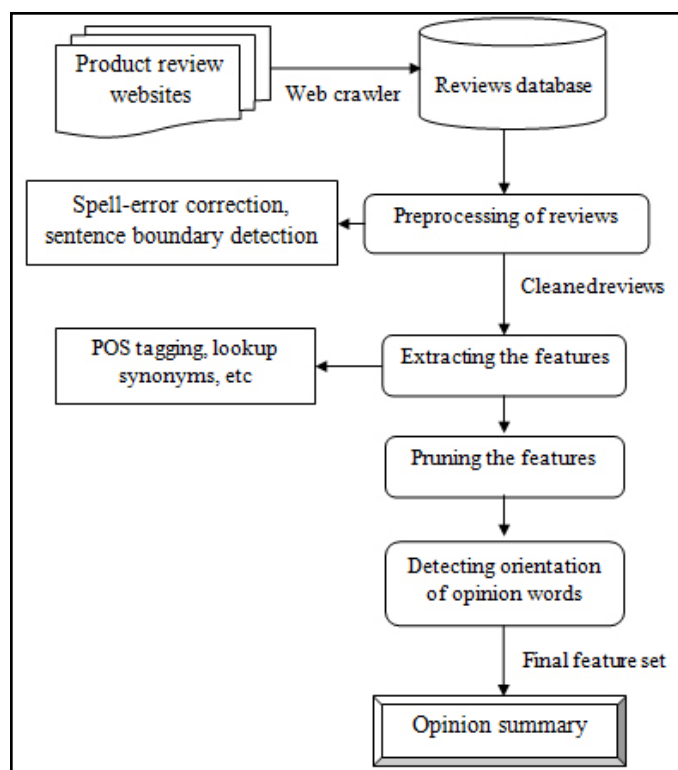


Fig. 1.1 : Opinion summarization system

If the customer wants more specifications on reviews, then the topic-based summaries or textual summaries are provided as the useful information for those customers. Irrespective of the summary formats, the main objective of opinion summarization is to provide the users for handling the huge availability of opinions

easily. The approaches utilized to address this summarization task are performed in different areas such as text clustering, sentiment prediction, text mining, Natural Language Processing (NLP) analysis, and so on. Some of these approaches executed based on simple heuristics, whereas others use robust statistical models.

2. Feature-Based Opinion Mining

Evaluative texts (e.g., a product review) classification based on the document level or the sentence level does not help for the opinion holder about their likes and dislikes. If there is a positive document on an object, then it is not to be said that the opinion holders have positive opinions on all characteristics or features of the object. Similarly if there is a negative document, it is not to be said that the opinion holder dislikes all about the object. In an evaluative document, the opinion holder provides the review about both positives and negatives of the object whether the general sentiment on the object may be positive or negative. In order to obtain such detailed positive or negative results, the feature level is required where three key mining tasks are employed as follows.

1.2.1. Identifying object features

Let us consider the sentence "The picture quality of this camera is wonderful" where the object feature is "picture quality". Supervised pattern mining method and unsupervised method are used for this problem. Here, frequent nouns and noun phrases are discovered as features, which are known as genuine features. For identifying object features, many information extraction techniques are provided such as conditional random fields (CRF), hidden Markov models (HMM), and many others.

1.2.2. Opinion mining determination

The features from opinion mining are positive, negative or neutral features for determining the task. A lexicon-based approach is developed in opinion mining for performing the determination of sentences level during feature extraction process. In addition to that, supervised learning approach is included for determining the direction of an opinion on features. Finally grouping operation is carried out with different features to complete the task on customer reviews.

1.3. Sentiment Analysis and Opinion Mining

A natural language processing is performed in sentiment analysis for following the features given in different products. Opinion mining is also known as sentiment analysis that produces the command, review and sentences of features. There are a number of important segments in sentiment analysis for providing feature extraction. They are sentiment classification, feature based sentiment analysis and opinion summarization. The main process of sentiment classification is to handle the features for classification purpose to the opinions against positive objects. Gaussian Probability Distribution function classification in the opinions mining attains better feature classification of certain objects. Nature language processing techniques (NLP) is used in this area for the function of sentiment detection in reviews.

1.4. Gaussian Probability Distribution function

Gaussian Probability Distribution function extracts the features from different customer reviews that are contained with various commands. The membership degree of a customer is estimated

by the Gaussian probability density function. The information associated with review commands are classified using Bayes classifier for better feature extraction. Examples include distinguishing different reviews given for hotels, online shopping, various brands for a single product, etc. These tasks yield a great deal of information for a customer and the user for determining better results on features extraction on customer reviews.

II. Literature Review

2.1. Opinion mining based on fuzzy domain ontology and Support Vector Machine: A proposal to automate online review classification

A Fuzzy Domain Ontology and Support Vector Machine system was explained in [1] based opinion mining system to computerize an online review classification. This paper presents a review on collection of information about hotel and hotel features. The customer feature reviews are identified by using SVM system and unrelated reviews are filter out with the help of Fuzzy Domain Ontology. The FDO and SVM are combined for improving the review rate and feature extraction of opinion mining. The proposed system significantly compute individual feature and provide the information in the form of text and sentiment analysis. Opinion features are extracted from review statement that are in opinion words and classified according to their mining features.

Initially, pre-processing stage collects the hotel reviews and classifier is used to classify the features for removing unrelated reviews. Next, a linearly separable hyperplane with binary categorization is used in SVM to classify the features reviews. This classifies into a positive reviews and negative reviews. Finally, all features are collected to form a hotel results on various reviews.

2.2. Extracting features from online software reviews to aid requirements reuse

A semi-automated approach known as Feature Extraction for Reuse of Natural Language requirements (FENL) were presented in [2]. The software features are extracted from the phrases by developing the software requirement specification along with requirements reuse. This approach composed of four stages for the extraction of software based on several combinations. Initially, the requirements of software reviews are accessed using open source service and they are extracted by online learning software. Next, similar features are extracted from attained reviews and special characters are removed for investigating the different features in collected documents. After identifying the features, related documents are considered and provide the pattern for the review extraction along with functional requirements. Finally, grouping process is carried out for computing the stored features together with review words. The feature grouping phase demonstrates the grouping process on extraction purpose in proposed semi-automated approach for providing an extracted features.

2.3. The Role of Emotions for the Perceived Usefulness in Online Customer Reviews

In [3], an extracting approach was presented from online reviews by various proportions with different products. The classification model is designed based on the review quantity and measures customer rating. The identification of feature characterize provides experimental evaluation with emotion dimensions and detect the categories of product. The crowd-funded solution consists of larger text features by online customer reviews and then performs

classification for obtaining the better online review commands. The suggestion for analyzing the online reviews is performed in four processes. Initially, they analyze the online customer reviews and then next it develops the model for identification of review without outer support. Next, it constructs the model and finally, it prefer the phrase book to presents an extraction model.

2.4. Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance

In this paper [4], Intrinsic and Extrinsic Domain Relevance (IEDR) approach was proposed to identify opinion features from online reviews. There are two differences in opinion feature statistics such as, one domain-specific corpus and one domain-independent corpus. The characteristic features from online reviews are collected in the form of text and extracted by syntactic dependence rules. The customer feature extraction is approximated with intrinsic-domain relevance and extrinsic-domain relevance. The proposed approach established the identification of opinion features that are presented in sentence level opinion mining. Based on distribution and difference, domain relevance is grouped together and measures the significant terms by the characterized domain relevance. The distributional significance of different document in horizontal significance is given by the Dispersion approach. Similarly, distributional significance of different document in vertical significance is given by the Deviation approach.

2.5. Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach

A new framework was developed in [5] by the combination of sentiment analysis and VIKOR (Vise Kriterijumsa Optimizacija I Kompromisno Resenje) approach. The multi conditions are provided in proposed approach for the satisfaction of mobile services that groups maximum utility of review. Data collection with pre-processing and customer satisfaction are main two process involved in proposed framework. Initially, text mining is employed to accumulate customer-review-based vocabulary of attributes and sentiment words during data collection and pre-processing stages. Next, customer satisfaction is considered by conducting the customer reviews of mobile application services.

The attributes and sentiment words are compiled using text mining on the basis of collected data from various customers' reviews. The review data is transformed into customer vectors with their references and each attribute is comprised with sentiment rating scores to form a keyword vector. According to the different mobile services, keyword vectors are included for the measurement of customer satisfaction stage. With the help of measured vectors, mass of each attribute is considered. The customer satisfaction information is provided with the help of maximum group utility and individual regret for identification of service operation.

2.6. Semantic-Based Unsupervised Hybrid Technique for Opinion Targets Extraction from Unstructured Reviews

The optimization of unsupervised Likelihood Ratio Test (LRT) technique was designed in [6] for opinion targets identification from unstructured reviews. The benefits of unsupervised technique are domain independent and it does not require any external training data. Based on the observed features, likelihood ratio test approach is proposed along with infrequent features. The proposed approach develops Word Net lexical resource to categorize infrequent features based on semantic features. The

review document of target opinion improves the accuracy of features extraction algorithms.

The proposed technique for opinion target identification is designed with three process namely, pre-processing, selection of candidate and opinion target selection. Initially, pre-processing is performed for removing the noise and part-of-speech tagging (POS). Here, imperfect sentences and unidentified words are removed during noise removal process. Next, noun phrase patterns are used for the selection of candidate features with identification of respective features. Finally selected features are classified with semantic-based similarity.

2.7. Feature Extraction and Analysis of Online Reviews for the Recommendation of Books using Opinion Mining Technique

The feature extraction and analyzing technique [7] classify the features of books from the reviews of the customers. The extracted features make a decision for readers to suggest the book that provides purchasing during online shopping. The customer's reviews replicate the customer's sentiments and have a significant products individual sell using online. Hence, the exacted features from the product analyses the text of reviews that involves a group of efforts and human intelligence. The requirement of the user is satisfied using feature extraction and analyzing technique and provides categorization method by users for indicates the features of books.

2.8. EXPRS: An extended pagerank method for product feature extraction from online consumer reviews

The customer reviews are obtained with the product of customer information and product uncertainty. An extended Pagerank algorithm [8] was proposed to integrate the contained feature with customer review expansions for producing the features automatically. The three different products are used for experimental estimation that shows more effective results on feature extraction from online consumer reviews. The proposed algorithm removes product features from a review corpus with the relation of nouns/noun phrases and sentiment terms.

Different features are identified and analyzed for generating candidate features and then filtering is applied for removing unwanted noise features. Rule based filtering is involved to eliminate non-feature nouns or noun phrases in proposed methods for producing customer reviews effectively. All positive and negative reviews are collected from customers for providing specific features extractions.

2.9. MapReduce Functions to Analyze Sentiment Information from Social Big Data

The paper [9] proposes Hadoop Distributed File System (HDFS) using MapReduce functions for sentiment analysis. Opinion mining extracts information for opinion mining of social multimedia data with meaningful information. The proposed method removes sentiment information from different category of unstructured social media text data from social networks to keep social multimedia data and using MapReduce functions for sentiment analysis. The data gathering and data loading is performed for maintaining the stable load of memory and CPU resources during data processing. The sentiment analysis is carried out using MapReduce functions for providing reduced feature extraction.

2.10. Extracting Implicit Features in Online Customer Reviews for Opinion Mining

A novel co-occurrence association based implicit feature extraction method [10] was proposed to review the opinion information for the customers. Opinion mining is used for feature extraction for providing fine grained mining results. The number of customer reviews is improved by developing e-commerce. User reviews are attained with customer information and feature transactions. Therefore, feature-based opinion mining results are provided with customers purchase decisions and producers to encourage their business.

2.11. Extracting Product Features and Opinion Words Using Pattern Knowledge in Customer Reviews

Extracting product features and opinion words was developed in [11] to provide an efficient feature extraction using pattern knowledge in customer reviews. The review features are produced in text format by obtaining the opinion word patterns and expressed using commands for customer. The information features are extracted with significant patterns of opinion mining. The reviews from customer express their opinion with phrases or words with different efficient patterns. Pattern extraction approach comments the customer review that gives the information about services and online products. Feature based opinion generates the customer review results for the determination of customer opinion that provides both positive and negative commands given in reviews.

Opinion summarization is provided by feature extraction and opinion word extraction for performing extraction process. Here, part-of-speech tagging is executed for identifying the feature reviews with nouns/noun phrases. Then nearest opinion words are extracted using feature extraction process for selecting the review sentences. Therefore, features are extracted with opinion-oriented words from customer review text.

2.12. A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data

This paper [12] has presented with high-dimensional microarray cancer data to reduce the dimensionality for analysis review results. Feature extraction method select the features for extraction with integrate preceding information from various natural sources. The computational complexity of data review is reduced by applying machine learning approach along with clustering methods. The size of data is increased with the application of hierarchical clustering based on different number of features.

Bi-clustering is preferred than the hierarchical clustering for extracting the feature information from the microarray data. After extraction of features, the size of data is minimized and develops original variables with selected features. Linear and nonlinear approaches are the two extensive groups for feature extraction algorithms. Here, lower dimensional data is extracted using linear feature extraction algorithm with the help of matrix factorization. Similarly, nonlinear extraction algorithm reduces the dimensional of features.

2.13. Critical Product Features Identification Using an Opinion Analyzer

Opinion mining system based on novel review and feature ranking methods were developed in [13] for identifying dangerous product features from enormous consumers' reviews. The internet facilitates the online customer reviews by developing the target

connections for generation of products. The features are determined with customer decisions and target analysis for improving feature ranking process. The reviews are ranked using ranking method that integrated with different number of features and opinion words. These ranked features are used for making the decision for selection of review customer commands. After selecting the reviews, higher review rate is preferred for feature ranking task with better opinion strength.

2.14. Enhancing Business Intelligence by Means of Suggestive Reviews

This paper [14] focuses on both consumer's choices and designers by means of suggestive reviews. Initially, the online reviews are identified and classified with business environment for satisfying the customer requirements. Next, different reviews are categorized into suggestive sentences either by explicit or implicit commands. Finally, sentiment analysis is performed on classified reviews with the applications of supervised machine learning approaches. Opinion mining extracts the features of customer review to achieve maximum performances of sentiment analysis.

2.15. Aspect-Object Alignment with Integer Linear Programming in Opinion Mining

A novel and important sentiment analysis task named as Integer Linear Programming (ILP) was proposed in [15] for obtaining accurate matching object in opinion mining. An aspect-object alignment classifier is developed for classifying the objects with different features for achieving target features. Here, intra-sentence constraints and inter-sentence constraints are imposed with ILP for interfacing procedure. Therefore, the feature sets attained from sentiment analysis performs better classification approach using an aspect-object alignment classifier.

2.16 Rule-based Methods for Extracting Product Features from Online Consumer Reviews

The rule based methods were developed in [16] for extracting (RubE) the product features from Online Consumer Reviews. RubE is an unsupervised rule-based technique used for extracting the subjective and objective features from online consumer reviews. The objective features are identified through integrating relation between part-whole and review-specific patterns. Then the subjective features are extracted by expanding double propagation with indirect dependency and relative structure. Therefore, the result of RubE method improves the significance performance than the state-of-the-art techniques.

2.17 Extracting Product Features and Opinion Words Using Pattern Knowledge in Customer Reviews

Due to the growth of e-commerce and web technology, the several online Merchant locates to write comments about the products for customer. Customer reviews expressed opinion about products or services which are considerably specified as customer feedback data. Opinion extraction about products from customer reviews is an important area of research and it is stimulated to improve an automatic opinion mining application for users. Therefore, Extracting Product Features and Opinion Words Using Pattern Knowledge are explained in [17]. For every feature, the opinion words are extracted from customer reviews. This method is used to obtain the patterns of opinion words regarding the feature of product from the review text about adjective, adverb, verb, and noun. The extracted features and opinions are valuable

for producing a significant review that can offer considerable informative resource to help the user and merchants to track the more appropriate option of product.

2.18. Intelligent Topical Sentiment Analysis for the Classification of E-Learners and Their Topics of Interest

An unsupervised algorithm called as Bigram Item Response Theory (BIRT) was introduced in [18] with the opinion polarity of all message related to the topic, author concentrate on sentence level opinion classification. The BIRT is performed with three processing level. Initially, lexicon based Approach is used to measure the sentiment through a number curving lexicons agreement. Bigram Cooccurrence relationship is determined using naïve Bayes Classifier. Finally, the item response theory is used in BIRT for relevant classification inference. The experimental results of the BIRT improve performance result than the other supervised algorithms.

2.19 Retrieving Product Features and Opinions from Customer Reviews

A language modeling approach was introduced in [19] using the aspect-based summarization task for retrieving product features from free-text customer reviews collection about the product. The natural language processing (NLP) techniques integrates a probabilistic model of opinion words and a stochastic mapping model. A novel method is used for the retrieval of product features from a group of customer reviews regarding a product or service. The proposed methodology failed to consider any training set of product features, and the experiments carried out over a few collections of customer reviews in English and Spanish.

2.20 A Review of Feature Extraction Software for Microarray Gene Expression Data

Feature extraction [20] is defined as changing large-scale gene expression data into a set of genes. A review of much software function is used for feature extraction. The software reviewed is mostly used for Principal Component Analysis (PCA), Independent Component Analysis (ICA), Partial Least Squares (PLS), and Local Linear Embedding (LLE). These software applications include a few restrictions about the statistical features and computational performance.

III. Gaussian Probability Distribution Based Feature Extraction from Customer Reviews

Opinion Mining is most commonly used to extract the public opinion from web contents about different domain. Web Content Mining obtains the user information known as sentiment analysis from various user opinions. Opinion mining detects the customer reviews in the form of text with the help of computational technique. Collections of information given in reviewer commands are presented in text document and analyze the customer review for extraction the features about product. Due to the rapid development of e-commerce, blogs and forums, mining opinions on the web is significantly important for feature extraction.

Initially, some of the customer review commands from opinion words are stored with their extracted features for Opinion Mining. All the opinion words are extracted based on positive and negative features and they are compared with stored mining features. The opinion mining with Gaussian distribution problem has been broadly divided into the collection of review, identification of

features, classification of like and dislikes features and opinion summarization. The feature identification in a sentence or document presented in customer review is capable to identify evaluative expression. Also some features are not explicitly presented and some are predicted from terms semantics called implicit features.

Naive Bayes classification denotes the relationship between distributed features from customer for opinion mining. Gaussian Probability Distribution for Feature Extraction is based on detecting and extracting the features into particular objects. Gaussian Feature Extraction obtains different customer reviews as input using opinion mining at the same time period and extracts the features. The feature extraction is performed and classified by using Bayesian classifier.

3.1. System Architecture

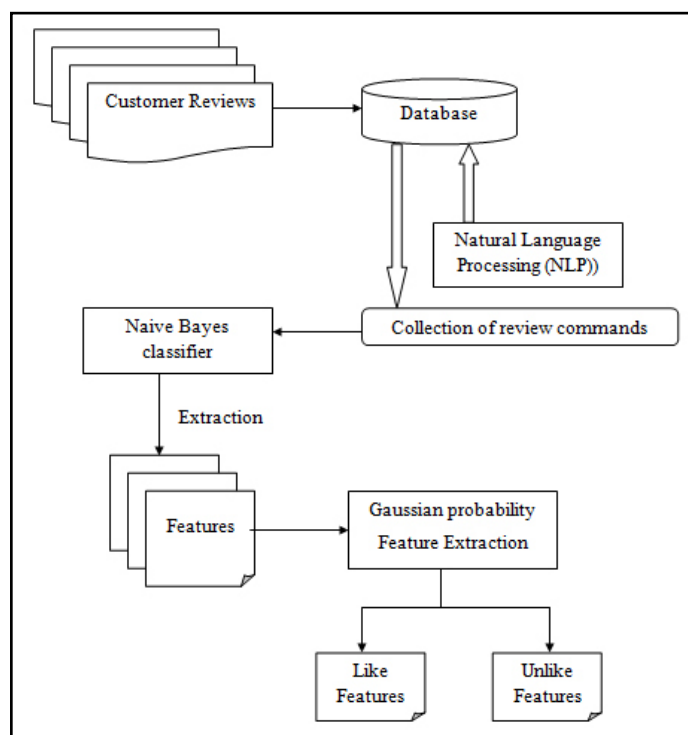


Fig. 3.1 : System architecture diagram of Gaussian Probability Distribution based Feature Extraction for Opinion Mining from Customer Reviews

Above figure 3.1 provides the system architecture outline extraction of customer review features. The opinion feature determines the summarization into two steps such as, feature extraction and opinion direction identification. The product name and an entry page for all evaluation of product into input system. Finally, result of evaluation is shown in the extraction process. The process of Gaussian Probability Distribution based Feature Extraction for Opinion Mining from Customer Reviews is divided into following steps.

- Customer Review Data for Opinion Mining
- Natural Language processing
- Naive Bayes Classification and summarization technique
- Gaussian probability Feature Extraction

a) Customer Review Data for Opinion Mining

The database illustrate in opining mining consists co-occurrence

relation data that are presented among different opinion words. The customers express their opinions or reviews on different product features in a single review. Co-occurrence relative patterns among features and opinion words are measured and quantified. They utilize distributional characteristics of opinion features in review corpus to identify the review commands. There are three different terms provided by opinion mining for attaining the customer reviews such as feature-opinion, feature-feature, and opinion-opinion.

Initial process for opinion mining is pre-processing step that collects the customer review from various review pages. The collected data consists of different labels and review documents are cleaned for removing the labels then performs the extraction process. The extracted reviews are stored in database for collecting the customer reviews at the sentence level.

b) Natural Language processing

The significant features are attained with Natural Language Processing (NLP) for extracting the features as likes and dislikes given by the customer reviews. The unstructured review commands on repeated iterations uses NLP for asymptotic assumption of opinion and feature words from the document. Each sentence given by the document provides part-of-speech for identifying the customer features. The relation between phrases in the topic is related to the document based on features or opinions. NLP task obtain the thoughts of customer that are expressed in positive or negative review comments by analyzing a large number of documents.

c) Naive Bayes Classification and summarization technique

Naive Bayes classification technique is processed with the help of association rule for the identification frequent features that are occurred in customer reviews. A set of opinion words or phrase is located as noun phrases and non noun phrases for the selection candidate features. Relevance features are used to classify the features into like and dislike feature. Transaction files are used for extracting the information that is occurred frequently in a stored dataset. Association rule is based on apriori algorithm where the frequent datasets are discovered for classifying the features. Bayesian approach suggest the customer review data based on opinion mining that extract multi product reviews and classify them according to the stored dataset. The feature extraction with opinion classifier uses the opinion task for ranking the features automatically.

d) Gaussian probability Feature Extraction

Gaussian probability Feature Extraction for opinion mining from customer reviews. The feature extraction from various customer reviews consists of different opinion about products, positive and negative reviews, infrequent features and implicit features. A set of opinion words or customer reviews are extracted from the domain data and classified according to the stored dataset. The customer reviews are estimated from different features by using Gaussian probability density function. The process of Gaussian Probability Distribution (GPD) for Feature Extraction from customer reviews is given below:

Input : Number of customer reviews in database
Output: Feature extraction
Begin
Step 1: Perform preprocessing and Collects all the customer reviews from the opinion words
Step 2:For each customer reviews in data base
Step 3: perform Natural Language Processing task to identify the relation between phrases in the topic
Step 4: Perform the classification using Naive Bayes classifier to determine the frequent features occurred in customer reviews
Step 5: Measure the Gaussian probability density function for extracting the different features from the domain data
Step 6: End for
Step 7: End

IV. Performance Metrics

The performance analysis is carried out in this paper with the metrics of True Positive Rate, Feature Extraction Accuracy and Feature Extraction Time. The performance metric evaluates and analyzes the values in java environment. Following metrics are used for experimental purposes.

- i. True Positive Rate
- ii. Feature Extraction Accuracy
- iii. Feature Extraction Time

4.1. True Positive Rate

The True positive rate is defined as the detection of positive reviews given by the customers on each features of the objects. It measures the proportion of positive thoughts successively according to the different number of documents that provides customer review commands. True positive rate is calculated in percentage (%).

Table 4.1 : Tabulation of true positive rate (%)

Number of records	True Positive Rate (%)	
	Existing IFOM	Proposed GPD
5	60.13	75.81
10	61.42	77.32
15	63.23	79.12
20	64.13	80.45
25	69.23	84.81

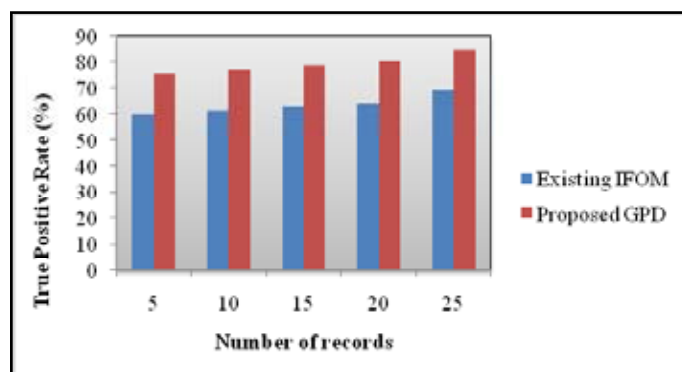


Fig. 4.1 : Measure of true positive rate

Above figure demonstrates true positive rate with respect to different number of records ranges from 5 to 25 that are obtained from customers. Figure shows the comparison of existing Identify feature opinion mining (IFOM) and proposed Gaussian Probability Distribution (GPD). When the number of records increased, true positive rate is also get increased. Therefore, Gaussian Probability Distribution function achieves higher true positive rate of 25% when compared to the existing Identify feature opinion mining approach.

4.2. Feature extraction accuracy

The feature extraction accuracy is defined as the measure of number of features that are correctly extracted from the customer reviews. Feature extraction is regarded with number of records given by the customers. The feature extraction accuracy is measured in terms of percentage (%).

Table 4.2 : Tabulation of Feature extraction accuracy (%)

Number of records	Feature extraction accuracy (%)	
	Existing IFOM	Proposed GPD
5	68.12	76.48
10	69.84	78.93
15	71.56	79.19
20	73.91	81.02
25	74.12	83.06

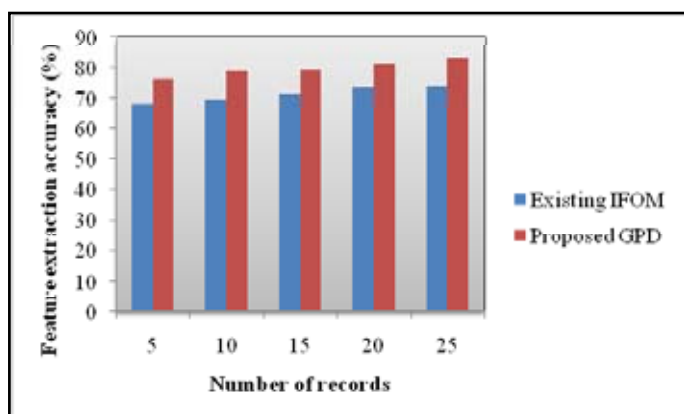


Fig. 4.2 : Measure of Feature extraction accuracy

Above figure shows the analysis of feature extraction accuracy with respect to different number of records ranges from 5 to 25 that are obtained from customers. Figure shows the comparison of existing Identify feature opinion mining (IFOM) and proposed Gaussian Probability Distribution (GPD). When the feature extraction accuracy is higher, the method is said to be more efficient. Therefore, Gaussian Probability Distribution function achieves higher feature extraction accuracy of 12% when compared to the existing Identify feature opinion mining approach.

4.3. Feature extraction time

The feature extraction time measures the time needed to classify all the features of the records from the customer review. In order to provide minimum feature extraction time, all the features are classified according to the stored datasets. It is measured in terms of milliseconds (ms). Lower feature extraction time ensures efficiency of the method.

Table 4.3 Tabulation of Feature extraction time (ms)

Number of records	Feature extraction time (ms)	
	Existing IFOM	Proposed GPD
5	13.4	8.5
10	14.6	9.7
15	16.2	11.3
20	17.8	12.47
25	19.1	14.5

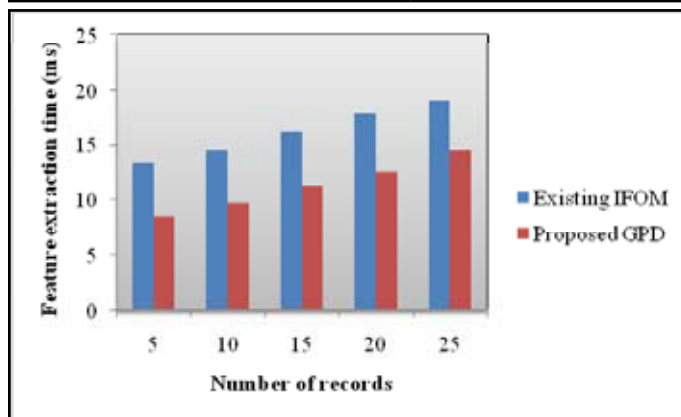


Fig. 4.3 : Measure of Feature extraction time

Above figure shows the analysis of feature extraction time with respect to different number of records ranges from 5 to 25 that are obtained from customers. Figure shows the comparison of existing Identify feature opinion mining (IFOM) and proposed Gaussian Probability Distribution (GPD). Therefore, Gaussian Probability Distribution function achieves higher feature extraction time of 31% when compared to the existing Identify feature opinion mining approach.

V. Conclusion and Future Work

Finally, Gaussian Probability Distribution (GPD) function provides better identification of Feature Extraction for Opinion Mining from Customer Reviews. The customer reviews are in the form of text, speech and database source that are distributed with various datasets. The duplicate reviews stored in the database are extracted using Natural Language Processing technique and it summarizes the customer review. The Gaussian Probability Distribution extracts the clear features at the sentence level of customer review and removing the unwanted features for opinion mining. Furthermore, Gaussian Probability Distribution function is developed with Discrete Gaussian function for extracting the features with more accurate results.

References

[1] Classification Farman Ali, Kyung-Sup Kwak, Yong-Gi Kim, "Opinion mining based on fuzzy domain ontology and Support Vector Machine: A proposal to automate online review classification", Elsevier, Applied Soft Computing, Volume 47, October 2016, Pages 235–250.

[2] Noor Hasrina Bakara, Zarinah M. Kasiruna, Norsaremah Sallehb, Hamid A. Jalab, "Extracting features from online software reviews to aid requirements reuse", Elsevier, Soft Computing, Volume 49, December 2016, Pages 1297–

- 1315.
- [3] Armin Felbermayr, Alexandros Nanopoulos, "The Role of Emotions for the Perceived Usefulness in Online Customer Reviews", Elsevier, *Journal of Interactive Marketing*, Volume 36, November 2016, Pages 60–76.
- [4] Zhen Hai, Kuiyu Chang, Jung-Jae Kim, and Christopher C. Yang, "Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance", *IEEE Transactions on Knowledge and Data Engineering*, Volume 26, Issue 3, March 2014, Pages 623–634.
- [5] Daekook Kang, Yongtae Park, "Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach", Elsevier, *Expert Systems with Applications*, Volume 41, Issue 4, March 2014, Pages 1041–1050.
- [6] Khairullah Khan, Baharum B. Baharudin, Aurangzeb Khan, "Semantic-Based Unsupervised Hybrid Technique for Opinion Targets Extraction from Unstructured Reviews", Springer, *Arabian Journal for Science and Engineering*, May 2014, Volume 39, Issue 5, pp 3681–3689.
- [7] Shahab Saquib Sohail, Jamshed Siddiqui, Rashid Ali, "Feature Extraction and Analysis of Online Reviews for the Recommendation of Books using Opinion Mining Technique", Elsevier, *Perspectives in Science*, Volume 8, September 2016, Pages 754–756.
- [8] Zhijun Yan, Meiming Xing, Dongsong Zhang, Baizhang Maa, "EXPRS: An extended pagerank method for product feature extraction from online consumer reviews", Elsevier, *Information & Management*, Volume 52, Issue 7, November 2015, Pages 850–858.
- [9] Ilkyu Ha, Bonghyun Back and Byoungchul Ahn, "MapReduce Functions to Analyze Sentiment Information from Social Big Data", Hindawi, *International Journal of Distributed Sensor Networks*, Volume 2015, Pages 1-11.
- [10] Yu Zhang, Weixiang Zhu, "Extracting Implicit Features in Online Customer Reviews for Opinion Mining", *International Conference on World Wide Web*, May 2013, Pages 103-104.
- [11] Su Su Htay and Khin Thidar Lynn, "Extracting Product Features and Opinion Words Using Pattern Knowledge in Customer Reviews", Hindawi, *The Scientific World Journal*, Volume 2013, Pages 1-5.
- [12] Zena M. Hira and Duncan F. Gillies, "A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data", Hindawi, *Advances in Bioinformatics*, Volume 2015, Pages 1-13.
- [13] Azra Shamim, Vimala Balakrishnan, Muhammad Tahir, and Muhammad Shiraz, "Critical Product Features Identification Using an Opinion Analyzer", Hindawi, *Scientific World Journal*, Volume 2014, Pages 1-9.
- [14] Atika Qazi, Ram Gopal Raj, Muhammad Tahir, Erik Cambria, and Karim Bux Shah Syed, "Enhancing Business Intelligence by Means of Suggestive Reviews", Hindawi, *Scientific World Journal*, Volume 2014, Pages 1-11.
- [15] Yanyan Zhao, Bing Qin, Ting Liu, Wei Yang, "Aspect-Object Alignment with Integer Linear Programming in Opinion Mining", *Plos one*, Research Article, May 2015, Pages 1-18.
- [16] Yin Kang, Lina Zhou, "RubE: Rule-based Methods for Extracting Product Features from Online Consumer Reviews", Elsevier, *Information & Management*, May 2016, Pages 1-41.
- [17] Su Su Htay and Khin Thidar Lynn, "Extracting Product Features and Opinion Words Using Pattern Knowledge in Customer Reviews", Hindawi, *The Scientific World Journal*, Volume 2013, Pages 1-5.
- [18] M. Ravichandran, G. Kulanthaivel and T. Chellatamilan, "Intelligent Topical Sentiment Analysis for the Classification of E-Learners and Their Topics of Interest", Hindawi, *Scientific World Journal*, Volume 2015, Pages 1-8.
- [19] Lisette García-Moya, Henry Anaya-Sánchez, and Rafael Berlanga-Llavori, "Retrieving Product Features and Opinions from Customer Reviews", *IEEE Intelligent Systems*, Volume 28, Issue 3, June 2013, Pages 19-27.
- [20] Ching Siang Tan, Wai Soon Ting, Mohd Saberi Mohamad, Weng Howe Chan, Safaai Deris, and Zuraini Ali Shah, "A Review of Feature Extraction Software for Microarray Gene Expression Data", Hindawi, *BioMed*, Volume 2014, Pages 1-15.