International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com Volume 4, Issue 1 (January-February, 2016), PP. 60-67

IDENTIFYING OPINION FEATURES USING INTRINSIC AND EXTRINSIC DOMAIN RELEVANCE

Aarti P. Nimbhore[1], Prof. S.B. Siledar[2]
MIT College of Engineering
aarti.nimbhore@gmail.com

Abstract— **There are number of techniques for opinion mining. But, mining the most frequent features from online reviews is the biggest task. If we use a single domain then it gives poor results. Using two or more domains is very difficult for feature mining process. In this paper, we proposed a novel technique for selecting and identifying most frequent features by using two similar or different domains (Main Domain and Sub Domain) which is based on two statistics.**

We first extract the features from Main domain by using some rules. For each feature (aspect) we then estimate its Domain Relevance Score which is called as Intrinsic Domain Relevance (IDR). Features those are having the less score (IDR score less than threshold) will be pruned and only those features get extracted which is having High IDR score.

In the final step we use our proposed EDR method. The features extracted from main domain having high IDR will be search out in sub domain. At the time of searching it will remove those features which are not available in sub domain. Hence, DR will be compute only for available features. Which known as EDR. Finally, get the most relevant features from sub domain or both the domains. Experimental results based on two different Worldreview domains show the improved EDR approach.

Keywords- **Opinion Mining, Opinion Feature Mining, Online Reviews, OpenNLP.**

I. Introduction

Opinion mining (also known as a Sentiment Analysis) aims to analyze people's opinions, sentiments, and attitudes toward entities such as products.Opinion Mining Process is illustrated in

Fig.1 [1].Generally individuals and companies are always interested in people opinions like if someone wants to purchase a new product, then firstly, he/she tries to know what other people think about the product and based on those reviews, he/she takes the decision.

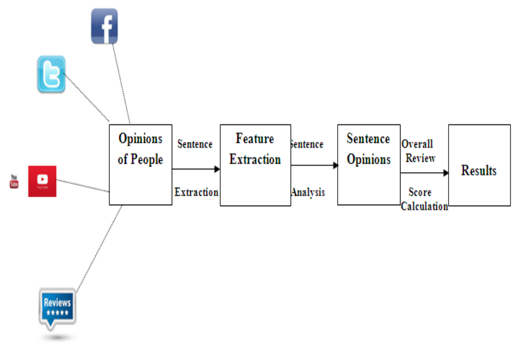


Figure 1: Opinion Mining Process

Opinion Mining means to track public mood. Itmeans to mine or extract a large amount of information and processing will be done on only important information after gathering them together. Hence, it is also called as Information Extraction system. Opinion mining follows some NLP concepts. Hence, it is a Natural language NLP.

Techniques like, computational linguistics and Information Retrieval (IR) combines and use for this mining process.

Opinions and sentiments expressed in text reviews can be generally analyzed at the document, sentence, or even phrase (word) levels.
The main objective of document-level (sentence-level) opinion mining is to classify the overall subjectivity or sentiment expressed in an individual review document [2].
It is difficult task to identify the opinion and relevant features on a particular product or service

For e.g. a mobile battery, mobile screen, mobile keypad etc) in order to make decision whether to purchase the service or product or not usually for task has been counted in sentiment analysis.

1. Opinion Identification
2. Feature mining
3. Classify opinion orientation
4. Summarizing Results

Feature extraction is most important and challenging task in opinion mining and analysis. To automatically extract the features it requires the Natural Language Processing (NLP) techniques. Following is the Example of sentence level opinion mining.

Example1.1. "The exterior of this mobile is very attractive; price is not expensive, though the battery backup is not good as compare to other phone. I still firmly like and recommend this cellphone."

From the above example, opinion on mobile which shows opinion orientation based on their features. Like, "exterior, price, mobile and cellphone" shows positive polarity and "battery" has negative orientation.

Availability of internet is increasing, due to internet, people can able to share and discuss their opinions with others across the worlds about particular topics. The social networking sites such as, Facebook, Twitter, Google plus, and other web forums are known as back bone for the online communication.Different scheme are available to find the opinion and emotion of the people. Sentiment may be defined as the opinion or emotion of the people and it may be positive or negative.

For example in the year 2012, the US president Obama speech, people twit fifty three thousand twits on the twitter in one minute. It is very complex task to extract the opinion of people from millions of comments by reading one by one. But sentiment analysis and opinion orientation have now made it possible to identify the sentence polarityand experiences of people from online reviews and large amount of texts. Similarly people are going to describe their opinions on the internet to strangers via the social networks and other web forums.

In the proposed work is based on feature extraction for two different domains. During identification and extraction many features extracted but few of them are important for consideration.

The proposed method extracts only important features by estimating its Domain Relevance Scores which are helpful in decision making process and ignores the irrelevant. The contribution of this paper is to extract relevant features during the process of sentiment analysis and applying proper part of speech tagging is performed by using OpenNLP.

Rule based method to getting the feature list and estimating DR score for the features which is extracted from Main Domain and then searches them in a Sub domain for selecting most important ones which are considered for both the domains.

II. RELATED WORK

Related work is based on two types of mining

1. Opinion Mining and Sentiment Analysis
2. Opinion Feature Selection and Extraction

A. Opinion Mining

Opinions and sentiments expressed in text reviews are analyzed at the document, sentence, and aspect levels. The main objective to these levelsof opinion mining is to classify the overall subjectivity or sentiment expressed in an individual review document (sentence) [3].

B. Types of Sentiment Analysis

The analysis can be done at three levels namely document level, sentence level and Feature level analysis [2].

C. Document Level Analysis

The task at this level is to classify whether a whole opinion document shows a positive or negative opinions For example, given a product review, the system which determines whether the opinion sentence expressed about that product is positive or negative. Once it get identified the polarity of opinion. Then the opinion sentences should be classified document wise. Hence, the concept called document level analysis.

D. Sentence Level Analysis

To classify or divide the reviews as positive, negative and neutral polarity is called as sentence level analysis.

Example ofopinion mining at document level

E. Input Documents

1. The Exterior of this mobile is very attractive, price is not expensive. Though the battery backup is not good as compare to other phone; I still like and recommend this cellphone.

F. Output Documents

Positive Documents:

2. The Exterior of this mobile is very attractive. Price is not expensive. I still like and recommend this cellphone.

Negative Documents:

1. Though the battery backup is not good as compare to other phone

G. Sentence Level Analysis

Both the document and the sentence level analysis do not show what exactly people liked and did not like. In case of Aspect level, it performs and analyzes finer-grained analysis. Aspect level also called feature level (feature-based opinion mining and summarization).It classifies sentences/documents as positive, negative or neutral polarity based on the features of those sentences/documents commonly known as aspect-level or feature level sentiment classification.

H. Approaches for Sentiment Analysis

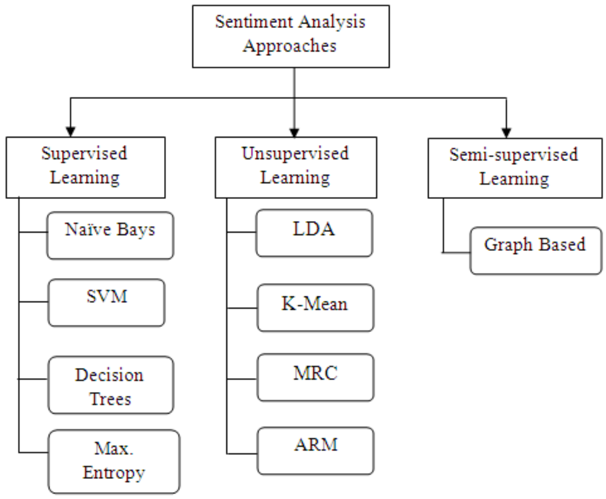


Figure 2: Approaches in Sentiment Analysis

Example

If you have to recognize which vehicle is Car and which one is Motorcycle?

Input data in supervised should be labeled. Hence, first you should assign that the vehicle which has 2 wheels and size is small is motorcycle.(Giving the information directly is supervised). But, in **unsupervised** learning, you can't label the inputs directly.You should give some different inputs to machine and the machine clusters with their similar features.

I. Supervised Learning Algorithms

The inputs are available to a machine learning algorithm is fully labeled. That means, all examples of classification that the machine is meant to reproduce. For this, a classifier is learned from the data, the process of giving input to the unseen case is called classification.Supervised learning is when the data you feed your algorithm is "tagged" to help your logic make decisions.There are some approaches used in *Opinion Mining* like, Supervised (Algorithms of classification) and Unsupervised Learning Models like clustering algorithms [4], [5].

Supervised learning models are including hidden Markov models [6]. Supervised models perform well on a given domain, but it need to training again when it applied to a different domain, [4], [5]. For a wide variety of products and services in different domains, supervised methods are less efficient because it is very expensive to construct assigned inputs foreach product or service (corpus). In addition, this model requires a decent sized set of assigned data for model learning on every domain.Supervised learning models that require given data have been successfully used to build sentiment classifiers for a given domain. In supervised learning model there are some problems regarding domain adaption.

Some of the most predominant Supervised Learning techniques[5], [7] in Sentiment Analysis have been SVM, Naive Bayesian Classifiers and other Decision Trees. Let us discuss one by one.

J. Naïve Bays Classifier

Naïve Bayes algorithm is implemented to calculate the probability of a data to be negative or positive.This method is based on the probability statement that was given by Bayes. This classifier provides conditional probability of occurrence of event E_1 when E_2has occurred already, the vice versa can also be computed by following mathematical statement[5], [7].

$$P(E1|E2) = \frac{P(E2|E1)P(E1)}{P(E2)}$$

This basically helps in deciding the polarity of data in which opinions can be classified as positive or negative polarity which is facilitated by collection of positive or negative reviews examples already fed.

K. SVM

The basic goal of support vector machine is to search a decision boundary between two classes i.e., excellently far away from any point in the training data.SVM develop hyper planes. Thesehyper planes in infinite dimension space. This distance from decision surface to closest data point determines the margin of classifier. So, the hyper planes act as decision surface which act as criteria to decide the distance of any data point from it. The margin is calculated by the distance from the closest data point. This success-fully creates a classification but a slight error does not cause any misclassification [5], [8].

L. Multi-Layer Perceptron (MLP)

MLP technique is quite popular owing to the fact that it can act as universal function approximate. A "back propagation" network has minimum one hidden layer with many non-linear units that can learn any function or relationship between set of input variables for discrete andconstant while output variable for discrete and constant. Due to this the technique of MLP is quite general, flexible and non-linear tools.

III. Decision Tree

A tree in which internal nodes are indicated by features, edges represent tests to be done at feature weights and leaf nodes indicate categories which results from above tests. It classifies a document by starting at the tree root and moving effectively downward via the branches (whose conditions are satisfied by the document) until a leaf node is reached. The document is then divided in the category that labels the leaf node. Decision Trees have been useful in many applications in speech and language processing.

A. Unsupervised Learning Algorithm

An unsupervised learning model was developed to classify review documents as thumbs up (positive) or thumbs down (negative) in the sentiment of each review document is predicted by the average sentiment polarity of phrases in the review. Domain-dependent contextual information is also considered for better estimation of the phrase sentiments. One drawback of this work is its reliance on an external search engine [5] [8].

Zhang et al.[8]proposed a rule-based semantic analysis approach to classify sentiments for text reviews. He used word dependency structure to classify the sentences of reviews, and predicted document-level sentiments via aggregating the sentence sentiments

In addition,Maas et al.[3],[9] proposed an approach to document-level and sentence level sentiment classification tasks, which uses a combine of unsupervised and supervised techniques to learn word vectors by capturing semantic term document information as well as loaded sentiment content. Clustering Classifier is an unsupervised learning technique which is used in Opinion Mining Purpose and there are some other unsupervised learning algorithms that are used in Opinion Feature Extraction and identification purpose only. Some algorithms are described below.

B. Association Rule Mining

Unsupervised statistics approaches use the results of statistical analysis on a given corpus or service to understand the distributional characteristics of sentiment features. The models are somewhat conflict to the informal nature of online reviews given a properly large reviews corpus.

Hu and Liu [10] proposed an association rule mining (ARM) algorithm to mine frequent features as potential opinion features, which are nouns and noun phrases with high sentence level frequency. However, ARM, which is based on the frequency of item sets, has the following drawbacks for the task of feature identification, 1) frequent but invalid features are extracted wrongly, and 2) rare but applicable features may be overlooked.

C. Mutual Reinforcement Clustering

To overcome feature based opinion mining problems, Suet al. [3] introduced a mutual reinforcement clustering (MRC) algorithm to mine the relation between feature categories and set of opinion words, based on a co-occurrence weight matrix generated from the given review product. Unlike several other corpus statistics methods, MRC extracts infrequent features, which provides that the mutual relation between feature and opinion groups found during the clustering phase is accurate. However, precision of MRC's is low due to the complexity in obtaining good clusters on real-life reviews.

D. Aspect Ranking Algorithm

Yu et al. [11] proposed an aspect ranking algorithm based on the probabilistic regression model to identify important or valid product features from online reviews. Moreover, their focus is not on extracting features commented on explicitly in opinions, but rather on ranking product features that are actually coarse grained clusters of specific features.

An Unsupervised learning modelof NLP extract opinion features by mining syntactic patterns of features in reviewsentences. In particular, the approaches attempted to discover syntactic relations among features and opinion wordsin sentences by using some syntactic rules or semantic role labeling. Syntactic relations [3]identified by the methods help to locate features associated with opinion words, itcould also extractlarge number of invalid features due to their colloquial nature of online reviews.Different Unsupervised topic modeling approaches, such as Latent Dirichlet Allocation (LDA), which have been used to solve aspect-based opinion mining tasks [3]. The models are developed primarily for mining latent topics or features, which is actually correspond to typicalconcepts of the expressed attributes, and may not necessarily be opinion features expressed explicitly in reviews. For example, LDA topic concept associated with cellphone reviews such as, "where" could be valid to this concept.Since some users want to talk about cellphone manufacturers, but it is not a specific opinionated cellphone feature. Therefore, though the approaches are successful in discovering latent structures of review data, they can't be much successful in dealing with identifying specific feature terms expressed explicitly in reviews.

E. Opinion Feature Mining

Opinion Featureis an entity on which user can expresses his/her opinions. Most of work has been done in feature extraction and identifying process. Somefeature mining approaches given below[10].

1. ARM
2. Feature Extraction algorithm
3. IEDR algorithm

The main reason for using **Association Rule Mining**[10] is because of the following observation. It is common that a online review of product contains unrelated data or information on which people can't identify feature of product. Most of people have different storiesevery time at the time of discussion.

However, whenever they comment about any product features on which discussion started, the words that they use come together. By using association rule mining algorithm for extraction of frequentitem sets is appropriate because those frequent item sets are considered as product features. Those features having noun/noun phrases tags which are not talked by people are to be considered as non-product features.

The main objective of**Feature Extraction**[12]is to select only important features and eliminate the less important ones. Important features are identified by computing two measurement values, these values generally known as statistics values. It is performed to keep those words which have highest score according to the predetermined measures using term frequency (TF).by the concept of term frequency it is assume important features occur more frequently than unimportant ones in the feature list.

$$tf(t,d) = \frac{f(t,d)}{\max\{f(w,d): w \in d\}}$$

Here, Tf (*t*,*d*) is identified as term frequency in a text documents. Which is defined as ratio of occurrence of term in a document to the maximum words arecontained in that particular text document.

The Existing approach IEDR, IDR and EDR is also based on selecting frequent features only. IEDR and EDR works for two different domains known as specific domain and independent domain. Its domain relevance with respect to the main domain and domain independent corpora is computed, which we termed as the intrinsic-domain relevance (IDR) score, and the extrinsic-domain relevance (EDR) score, respectively [3], [13]. But, existing EDR and IEDR has been extracted too many irrelevant features. So, to overcome this problem the new improved EDR approach has been implemented.

Proposed approach Improved EDR is based on IDR, first IDR extracts opinion features from main domainby applying some rules.In the second step we calculate DR. and to identify opinion feature we compare DR of each feature to its threshold value. Features those having fewer scoreare pruned out.In the third step improved EDR search that opinion features in the sub domain and calculate their measures. And in the last step we pruned those features those having less score.

IV. METHODOLOGY

A. Overview

Fig. 3 shows the workflow of proposed method. Proposed method is based on two different corpora. It can be also work for similar corpora. But given two different corpora like, domain dependent and domain independent.In the first step we extract a list of all dependent product features by using some dependencies grammar rules which is known as part of speech tagging [3], [13], [14]. In the next step we compute their scores which is called as domain relevance score. Generally we say it as IDR (intrinsic domain relevance). After that compare the scores with threshold values (threshold indicates valid opinion features) and extract only valid opinion features.
In the last step check the valid features in other domain for the availability of the features and again compute the scores, compare with the threshold and extract most valid features. At last the features those are extracted will be frequent features from both the domains so, we consider it as IEDR. But those are searched for independent corpus hence we can say it as improved EDR and for similar Domains we say improved IDR.

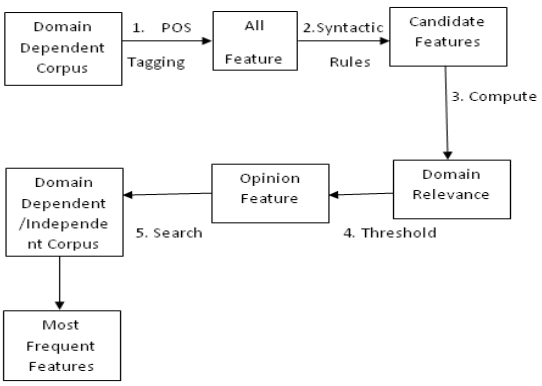


Fig. 3: Proposed Method Workflow

B. Opinion Feature Extraction

Product features are usually nouns or noun phrases in review sentences.Thus the part-of-speech tagging is crucial.

Part -of-Speech Tagging (POS)

We used the OpenNLPparser [15] to parse each review to split text into sentences to tokenize each word with their proper abbreviation with the help of part of speech tagging.After tagging each word will be identify with their proper tag like, noun, verb, adjective, etc. This process also identifies noun, proper noun, noun phrases and verb groupswhich we call it as syntactic chunking [3]. The following shows a sentence with POS tags [10], [12], [16].

The_DT price_NN of_IN the_DT Cellphone_NNP is_VBZ too_RB expensive_JJ.

After tagging, each sentence is store in the transaction database along with their POStag information for each word in the sentence. Each line contains words of every sentence,which consist only the identified nouns and noun phrases of the sentences. Other part of the sentence is unlikely to beproduct features. Removal of stop words includes in preprocessing technique. i.e., noise free data.
After Tagging all sentences applied some syntactic rules through which we can extract only Nouns (NN) and Noun phrases (NNP) as an opinion features.
From the above example we get the features **price and cellphone**after applying rules.

C. Opinion Feature Identification

Feature Identification based on Domain Relevance Score. It means how much a feature is related to a particular corpus. Two measures on which Domain Relevance Score is based i.e. Dispersion and Deviation [3], [13].
Dispersion indicates how significantly each feature is related across all documents in the corpus.
Deviation means how significantly the term mentioned in a each document in the corpus. Both the measures computed by using term frequency (*tf*), document frequency (*df*) and weight (*w*) of each feature. Here, document frequency is an inverse document term frequency (*IDTF*).
Following are some mathematical expressions on which we compute *DR*

D. Weight of Term

$$w_{ij} = \begin{cases} \left(1 + \log TF_{ij}\right) \times \log \dfrac{N}{DF_i} & if\, TF_{ij} > 0, \\ 0 & otherwise \end{cases}$$

Here, *wij* is a weight of term in each document in the corpus.*'i'* indicates term and *'j'* indicates document *'j'*.

E. Document Frequency

$$D_{ij} = \frac{TF_{ij} \times \log(No.of\ Documents\ in\ the\ corpus)}{\text{occurence of term}}$$

Here, *Dij* is a Document frequency of each term in each document in the corpus.Frequency indicates Occurrence of each term in each document in the corpus.

1. **Average weight $\overline{w}_j$ of all terms in each document D_j in the corpus**

$$\bar{w}_j = \frac{1}{M}\sum_{i=1}^{M} w_{ij}$$

2. **Deviation of term** ***Ti*****in the Document*****Dj***

$$devi_{ij} = w_{ij} - \bar{w}_j$$

3. **Average weight** $\bar{w}_i$ **of all terms across all documents** D_N **in the corpus**

$$\bar{w}_i = \frac{1}{2}\sum_{j=1}^{N} w_{ij}$$

Here, D_N is a total no. of Documents in the corpus.

4. **Standard Deviation** s_i **of term across all documents**

$$s_i = \sum_{j=1}^{N} \frac{(w_{ij} - \bar{w}_i)^2}{N}$$

5. **Dispersion**

$$disp_i = \frac{\bar{w}_i}{s_i}$$

6. **Domain Relevance** DR_i **of term**

$$dr_i = disp_i \times \sum devi_{ij}$$

The process for calculating the domain relevance is the same for both the domainsas well as IDR and EDR methods, as in Algorithm 1. When the procedure is applied to the domain specific review corpus, the scores are called IDR, otherwise they are called EDR.

Algorithm 1: Calculating Intrinsic Domain Relevance (IDR)
Input: A domain specific corpus C
Output: Domain relevance scores (IDR)

for each candidate feature CF_i **do**
for eachdocument $D_j$in the corpus *C*
do
Calculate Term frequency TF_{ij};
Calculate $DF_{ij}$by (2);
Calculate Weight $w_{ij}$by (1);
Calculate Average $\bar{w}_j$ by (3);
Calculate deviation $devi_{ij}$ by (4);
for all documents $D_N$in the corpus *C*
do
Calculate standard deviation $s_i$by (2);
Calculate dispersion $disp_i$ by (3);
Compute domain relevance dr_i by (5);
Return A list of domain relevance (IDR)
if ($idr_i \geq$ *ith)* ***then***
Return A list of opinion feature with IDR

Proposed EDR method is based on existing IDR method. First step extract features using IDR By applying algorithm 1. To extract opinion features compare DR with threshold value. Proposed EDR works on opinion features extracted from main domain. In the second step improved EDR search opinion features in sub domain by applying algorithm 2. The most important part in proposed is that we removed those features which are not available in the independent corpus.

Henceforth, at last get the list of most relevant features from both the domain but it's an EDR score of features. So, we say improved EDR. But, we can consider that features from both domains.

Algorithm2: Calculating Intrinsic and Improved Extrinsic Domain Relevance
Input: A domain specific corpus C and Independent corpus R
Output: Domain relevance scores (IDR/improved EDR)

for each candidate feature CF_i **do**
for eachdocument $D_j$in the corpus *C*
do
Calculate Term frequency TF_{ij};
Calculate $DF_{ij}$by (2);
Calculate Weight $w_{ij}$by (1);
Calculate Average $\bar{w}_j$ by (3);
Calculate deviation $devi_{ij}$ by (4);
for all documents $D_N$in the corpus *C*
do
Calculate standard deviation $s_i$by (2);
Calculate dispersion $disp_i$ by (3);
Compute domain relevance dr_i by (5);
Return A list of domain relevance (IDR)
if ($idr_i \geq$ *ith)* ***then***
Return A list of opinion feature with IDR
Then search the list of opinion feature in the other Independent Corpus R. while searching remove opinion feature which are not present in R.
Follow the same procedure for computing DR of each opinion feature in corpus R.
At last,
Return A list of domain relevance (improved EDR)

if ($edr_i \geq$ ***eth) then***
Return A list of opinion feature with improved EDR

4.2 Evaluation of Case Studies

If we take the two different corpora for e.g. cellphone (Dependent Review Corpus) and Electronic Devices or any other corpus (Independent) then we get the list of features those are extracted through **IDR and EDR.**

Example3.1."There is no flash on the camera but it takes very nice and clear pictures. The quality of call is nice"

Example shows a sample product reviews on cellphone. Applying Algorithm 1 on the above example as follows: First, apply the syntactic rules (Rules in short) to extract a list of candidate features (nouns): “camera,” “picture,” “call,”, “quality” and “flash”as shown in line 1 of Table 1. Next, filter the features by taking threshold value will get the features listin line 2 of table 1 (this is the existing IDR procedure).
Now apply the proposed algorithm 2 to get the list of features having EDR defined in line 3. Here, we get the features like, camera. Picture and call only.
This is because, whatever we get the list of features through existing IDR, those features we are searching in sub domain (independent domain) and those feature in the sub domain is not present that feature gets remove. So, the searching is only for those features which are available in the sub domain.

Table 1: Extracted Opinion Features via existing method and proposed method for two different corpora.

Rules	Camera	Picture	Call	Flash	Quality
IDR	camera	Picture	Call	Flash	
Improved EDR	**camera**	**Picture**	**Call**		

Here, in the abovetable, EDR in bold is our improved EDR.

V. EXPERIMENTAL RESULTS

A. Corpus Description:

Experimental Description is based on two different real word reviews which are taken from different social networking sites.

Table 2

Corpus	**Reviews**	**Sentences**
Cellphone	250	500
Electronic Devices	250	500

B. Results:

An experimental result is based on 4 measures. **Precision and Recall** is the basic measures used in evaluating search strategies [3], [5], [17].

1. **Precision**is the ratio of the number of relevant records retrieved to the total number of relevant and irrelevant records retrieved in the database.
2. **Recall** is the ratio of the number of relevant records retrieved to the total number of relevant items in the database. It is usually expressed as a percentage.
3. **Frequency Measurement**F-measure is a harmonic average of both precision and recall. Which is given below,

F-Measurement is equal to,

$$\frac{2 \times Precision \times Recall}{(Precision + Recall)}$$

4. **Accuracy**Here, in opinion feature mining the accuracy referred to the degree of a measure to standard relevant and irrelevant items in a dataset.

Accuracy is the portion of all relevant and irrelevant features against all features. An accuracy of 100% means that the features are exactly the same as the actual features. Precision is the portion relevant items against all features in the corpus or dataset. Recall is the portion irrelevant features against all actual features. F1 is a harmonic average of precision and recall [5]. Comparison among various parameters between existing and proposed method.

1. **For two different Review Corpora**

Domain Dependent Corpus- **Cellphone**
Domain Independent Corpus – **Electronic Devices**
Improved Methods: **EDR**

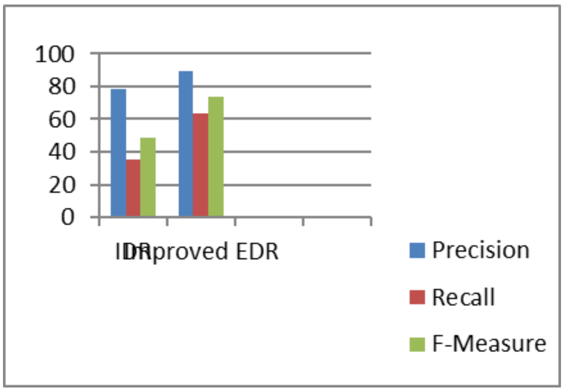
Table 3: Comparison among various parameters between existing and proposed method EDR.

Methods	**Precision**	**Recall**	**F-Measure**
IDR	78.22%	35.62%	48.95%
Improved EDR	**89.28%**	**63.29%**	**74.07%**

From the above table IDR is an existing method and the other in bold are proposed methods. Here, we compare our proposed method EDR with the existing IDR by taking some *ith threshold*for two different domains by taking *eth threshold.*

C. Precision VS Recall

1. **For two different Domains Mobile vs. Electronic devices**



VI. CONCLUSION

Improved EDR gives more relevant features by searching the candidate features in sub domain i.e. (domain independent) which extract the most valid features. Hence, a good quality domain-independent corpus is quite important.
The results have shown that proposed method of frequent feature extraction is efficient. Existing IDR extracts all the candidate features with irrelevant features. But, in case of an improved EDR, this method extracts frequent features which are relevant in both dependent and independent corpora.
The proposed work is tested with 500 cellphone reviews and got 60.28% accuracy for different Domains
As compare to Existing IDR, the improved EDR increases precision and recall by 11.06% and 27.67% for different Domain for 0.9 threshold.

References

[1] Pravesh Kumar Singh, "Analytical Study of Feature Extraction Techniques in Opinion Mining", VLSI-2013 pp. 85-94, 2013.
[2] Richa Sharma, "Opinion Mining of Movie Review at Document Level", Vol. 3, No. 3, July 2014.
[3] Zhen Hai, Kuiye Chang, Jung-Jae Kim, and Christopher C. Yang, "Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain", IEEE Transaction Vol. 26, No. 3, March 2014.
[4] Neha S. Joshi, "A Survey on Feature Level Sentiment analysis", Vol. 5(4), 2014.
[5] Jayshree Khairnar, "Machine Learning Algorithms for Opinion Mining and Sentiment Classification", Vol. 3, Issue 6, June 2013.
[6] A Novel Lexicalized HMM-based Learning Framework for Web Opinion Mining.
[7]Pravesh Kumar Singh, "Methodological study of Opinion Mining and Sentiment Analysis techniques", Vol. 5, No. 1, February 2014.
[8] A.V. Moholkar, "A Survey on Opinion Mining Techniques". ISSN, Vol. 2, October 2014.
[9] A.L. Maas, R.E. Daly, P.T. Pham, D. Huang, A.Y. Ng, and C. Potts, "Learning Word Vectors for Sentiment Analysis," Proc. 49th Ann. Meeting of the Assoc. for Computational Linguistics: Human Language Technologies, pp. 142-150, 2011.
[10] "Mining and Summarizing Customer Reviews".
[11] Zeng-Jun Zhang, "Aspect Ranking: Identifying Product Aspects from Online Consumer Reviews".
[12] Abdul Wahab, "Important features selection during sentiments analysis" 2014.
[13] Varsha Sarnikar, "Feature based analysis in opinion mining via Domain Relevance" ISSN, Vol. 1, Issue 3.
[14] L.J. Sankpal, "A Survey: Extracting Features from Online Reviews Corpus by Domain Relevance", Vol. 3, Nov. 2014.
[15]OpenNLP Parser.
[16]Grammar: Part 1, Part of Speech Tagging.
[17] Searching Recall and Precision.