Recommendations of Products based on Feature Based Sentiment Analysis and Demographic

M.Mallikarjuna Rao[1], Dr.K.Fayaz[2]

[1]Research Scholar, Rayalaseema University

[2]Faculty of Systems, SKIM, S.K.University,A.P., India

Abstract: ***The ecommerce system is using the entire modern way of shopping in the consumer market. There are lot of well established companies and new start ups which provide the ecommerce services. Lot of products coming into the market and are sold online with good strategies. On the other hand it is burden for the consumer to known the best products because most of the applications uses numerical rating scales with collaborative based filtering which does take into consideration the sentiments of the products expressed by the consumers online. In this paper feature based sentiment analysis with demography is presented which makes use of real time reviews collected using web crawler and then sentiment analysis is performed on the subset based on demographics the products is recommended for the end user. The approach is also compared with direct sentiment analysis with feature and demographic information. Also a search algorithm is provided which will rank the products based on positive sentiment descending, negative sentiment ascending and neutral sentiments descending if feature is present. if feature is not present then ranking based on highest frequency order is proposed.***

Keywords- Web Crawler, Tokenization, Frequency computation, Feature based Frequency, Review based sentiments. Product based sentiments, Ranking based frequency vector

I. INTRODUCTION

Ecommerce system is a process in which a merchant and a consumer which can be a direct consumer or another merchant itself exchange the information in order to perform an end to end transaction in which both the merchant and consumer have a system of interest. The nature of the ecommerce system and then the various fundamentals of ecommerce system along with the software tools are present in detail [1]. Day by day the numbers of users are increasing from thousands to millions who access the internet looking for products and other services. The ecommerce systems behave like a virtual market place instead of physical transactions. The factors which are leading the revolution are internet users and the technology.

There are many models available for the ecommerce system business to consumer (B2C) and business to business (B2B) in the market today. In B2C there is direct communication between the merchant and the direct buyer through online shopping. In B2B model one business in integrated with another business in order to complete the end to end delivery.

Few use cases of the recommendations can be the following

Collaborative based Recommendations

In Collaborative based recommendations [2] module the users which are unregistered or register each of the users will be able to rate the product based on the numerical rating. After that the aggregated sum is performed. Finally the ranking is performed based on maximum aggregated sum of rating.

Drawbacks of Collaborative Rating

[1] The entire rating is based on the numerical scale and it does not take into consideration the emotions or whether the products are quality products and lot of fraudulent ratings can also be given which will increase the advantage for a wrong product.

Content Based Recommendations

Content Based Recommendations [3] is user specific data in order to provide recommendations. In this module the user will select a product and enter the credit card details and then completes the transactions. Behind the scenes the merchant maintains the transactions and then finds the best products suited for the user.

Advantages

[1] The history profile of the transactions are used in order to recommend the products for the user.

Drawbacks

[1] In this approach the threshold is required

[2] The new users do not get any recommendations because there is no buying history for the user

Pearson Recommendations

This kind of recommendations [4] is given for the authenticated user. The ratings are taken from the user as well as other users and then based on prediction rating the products are recommended. The recommendations are computed by using the following equation

$$s(u,v) = \frac{\sum_{i\in I_u \cap I_v}(r_{u,i}-\overline{r_u})(r_{v,i}-\overline{r_v})}{\sqrt{\sum_{i\in I_u \cap I_v}(r_{u,i}-\overline{r_u})^2}\sqrt{\sum_{i\in I_u \cap I_v}(r_{v,i}-\overline{r_v})^2}}$$

Where

$r_u = rating\ \ from\ user\ u$

$r^{-}{}_u = average\ of\ \ all\ ratings$

$I_u = set\ of\ \ items\ rated\ by\ \ user$

The prediction is substituted using the following formula

$$p_{u,i} = \overline{r_u} + \frac{\sum_{u'\in N} s(u,u')(r_{u',i}-\overline{r_{u'}})}{\sum_{u'\in N}|s(u,u')|}$$

Sentiments based Recommendations

The sentiment based recommendations [5] take the reviews for the product and then computes the positive and negative of the product based on reviews and ranks the product. In this work the concentration is one sentiment analysis by bringing in two important modifications namely the feature set and then the demographic information.

II. BACKGROUND

There is a lot of work done in the literature with regards to the recommendation system. The personalized recommendation [6] process helps users to perform convenient and efficient shopping and also improve the efficiency of the system. Internet [7] is moving at a faster pace and daily tasks like online shopping, paying utility bills etc have changed the way of implementation based on new business needs. The audience has shifted to online shopping which provides the producers and retailers lot of customized options. The users perform the rating of application. Collaborative Filtering [8] predicts user preferences based on two techniques memory based and model based. Resources [9] in cloud computing like Amazon, Google app engine can be used for performing mobile search which can be both context aware and personalized activity. Hybrid filtering is used to eliminate irrelevant results based on combination of content and collaborative filtering. A recommendation system [10] finds the products by measuring the similarity of properties. A search engine is used which gives the query and then provides the results based on ranking of items based on positive preference. online recommendation [11] provides quicker way to buy items and complete transactions quickly. The recommendations system recommend products by using tools which has two important factors namely increase profits and retain buyers. Book recommendations can be performed by performing the intersection between the content based recommendations and collaborative based recommendations. The limitations of content and collaborative based filtering [12] can be minimized by using diverse item selection which gives dissimilar items and combined with content based to generate product based recommendations. Content based recommendations [13] provides the strategy which can find user preferences and compares the user preferred products with available products by providing enhanced personalized recommendations, computational viability and greater accuracy. The film recommender [14] systems filter the results based on actors, directors and genres. The system provides recommendations based on new and previous unrated movies. The maintainable information filtering system [15] provides simple and efficient solution which can block a list of IP Address. It is very difficult to block all IP Address. In order to overcome the problem URL filtering can be used multiple classification ripple-down rules (MCRDR) knowledge acquisition method allows domain expert to maintain knowledge and filtering system. A content-based filtering system [16] that targets music data in MIDI format. The approach performs the analysis of characteristics of feature parameters about music data in MIDI format. Twitter [17] is a popular social website which is used as a platform for expressing opinions and attitudes. Latent Dirichlet Allocation (LDA) based model can give potential interpretations of the sentiment variations and separate out longstanding background topics and then finally rank the tweets based on popularity. Predicting stock market [18] movements perfectly represents the public sentiment and opinion about current events. Dow Jones Industrial Average Index (DJIA) work is to observe the changes in stock prices of a company, the rise and falls of stock and then correlate with the public sentiments. Two different textual representations, Word2vec and N-gram are used then sentiment analysis and supervised machine learning principles are applied for tweets. If the tweets have positive news about the company which is an encouraging factor for people to invest in company. In a Multi domain environment [19] performs the estimation of polarity by using different domains. The linguistic overlap between domains provides a way to build a sentiment model by performing the inference of documents belonging to different domains. Data Analytics [20] is used to make industries and organization to improve business decision. The data are generally of two types namely structured and unstructured data. Opinion mining is used in a daily decision making process like mobile phone purchases and movie reviews. Bayes algorithm, Support Vector Machine, Maximum Entropy are the machine learning algorithms used for sentiment analysis which has only a limited sentiment classification category ranging between positive and negative. Sentiment Analysis [21] has been used to perform research using a method that can analyze different languages to find sentiments in them and perform sentiment analysis. The text is analyzed using machine translation techniques and then data is processed for finding the sentiments in the text. Internet [22] is very important part of modern life. OSN's provide a platform where people can share their views, ideas, information belonging to different locations. Sentiment analysis can be used to find positive and negative tweets using natural language processing and information extraction. Twitter [23] is used to voice public opinion. The challenge is used to extract specific results by keeping in mind the sentiments of public. Demonetization had negative and positive effects in various regions which are analyzed using sentiment analysis. The reviews [24] are used for performing the evaluation of various subjects across various products. Each product has various aspects named Mimetic Voter Patterns, or MVP, to identify aspect words, by using patterns of parts-of-speeches of their adjacent words. There is a voluminous amount [25] of opinionated data which is used to understand and exploit the business. An innovative approach is used for performing the analysis of textual data and find the positional and sentiment polarity. There are millions of reviews [26] for any product on the social media. There is a lot of time required by the user to go through those reviews and select the best product. The food recipes are ranked based on the sentiment analysis with positive as the maximum and negative as the minimum. There is lot of unstructured [27] text data and it is very challenging to find the sentiment analysis neural network architecture makes use of Convolution Neural Network (CNN) and Long Short-Term Memory (LSTM) along with the pre-trained word vectors. The internet [28] is responsible for publishing the news and the news gets divided into various categories such as politics, economics, sports, and health. Each area has a positive and negative sentiment. These sentiments are responsible for government policies and other interest of topics. Sentiment analysis can also be used to classify votes. The sentiment analysis and human agent interaction [29] performs sentiment related phenomenon along with detection and dialog management. In this approach sentiment analysis acts as input and then output is interaction strategies. The internet provides opinions [30] for investors. The sentiment analysis is used for investment decision making and risk perception. Sentiment ontology can be used to conduct context sensitive sentiments. There is a correlation between stock price and forum sentiment which increases accuracy. The flow of positive and negative news [31] is performed in the capital markets where assets are priced and risk assessed based on future expectations. The document level sentiments are then break down into fine grained levels with tweets and news acting as input and finally the sentiments are predicted. The derivation of lexica [32] priority is used to perform sentiment analysis based on positive and negative scores. A learning framework known as ensemble method is used to predict the sentiments using higher coverage. Opinion mining [33] is used for core tasking which makes use of Bag-Of-Words (BOW) for training a classifier in statistical machine learning. Dual Opinion mining Model is used which makes use of original and reversed review. There is a lot of rich data [34] available on the web in terms of blogs, review forums which offers a rich source of opinion data. Senti Lexical algorithm performs the numerical computation of positive, negative and neutral sentiments. The hybrid filter-wrapper [35] approach to sentiment polarity classification is a two-phase feature selection method. In the first step initial preprocessing is performed. The second step is to perform wrapper based feature selection which performs the integration of generic and support vector machines. The data is collected from IMDB website. Aspect base sentiment analysis [36] is a concept of machine learning. In this approach trained data is used to provide the positive, negative and neutral sentiments. Finally the total polarity is computed using Support Vector Machine (SVM) and Maximum Entropy (ME).Business need to find out the polarity [37] from the social media analytics and then perform smarter decisions. The same is applicable for politics in which the party or a candidate can change the strategy based on public opinion. Naive Bays and SVM are used to perform sentiment analysis on sentiment data. Turkish based sentiment lexicon [38] is used to perform analysis of positive, negative and neutral sentiments. The lexicon also increases the number of sentiment works by 10k. Arabic language has a nice vocabulary [39]. The unique properties of standard Arabic and dialects are used in order to perform supervised Arabic sentiment analysis using a bag-of-words feature. Twitter is a trusted platform for online micro texts [40] and is used for monitoring of public sentiments. Sentiment analysis is used to predict the public opinion for real world problems. Sentiment analysis [41] is used for performing calculative analysis of views, sentiments, opinions and positivity or negativity of a text. The algorithm identifies the factors responsible for different sentiments of a person. The computational treatment of opinion [42], sentiment, and subjectivity in text is important. A framework used for adding a sentiment sentence compression is responsible for making the sentence shorter. Twitter offers organizations [43] the ability to inspect public feeling on products and events. As part of first step there must be preprocessing. The preprocessing can be used to remove Uniform Resource Locator and stop words. The amount of information [44] has grown rapidly. Conventional recommendation systems have problem of scalability and inefficiency. The recommendation systems are based on static ratings and ranking of products for different users without specific needs. Personalized system is used to recommend items based on their needs. Recommendations system [45] is used to recommend items or products based on user interest and their ratings. Association rules are required to perform association among products. The amount of information overload [46] makes it very complex for users to get useful recommendations within the specified time. The recommender systems helps customer to make useful decisions to purchase products. The CF algorithms [47] and multitude parameters have impact of quality and improve classification accuracy of that particular algorithm

III. Proposed System and Analysis

The hybrid recommendations are the new innovation that is introduced newly. Although there are lot of work done on opinion mining all the sentiments are calculated at review level. In this thesis recommendations are given after computing polarity both at the review level and feature level. if the features are not present then no weight will be given for the positive and negative sentiments. The entire recommendations performs a sequence of steps namely Web Crawler Real Time Review Collections from top ecommerce sites using DOM and X-Path technology, Data Cleaning is performed in order to remove keywords known as stop words. After that tokenization, frequency computation and feature based frequency computation is performed. The lexical Positive and Negative keywords are taken and then review based negative and positive sentiments are computed. If both positive and negative sentiments are zero then neutral sentiments are set to 1. The process is performed per review and per feature. Finally a list of unique products are found out and for each of the product then the sentiments are computed by computing the positive, negative and neutral sentiments are computed per product and per feature. The following use cases are executed and recommendations are performed for the product in such a way that positive sentiment is highest, negative sentiment is lowest and neutral sentiment is maximum. For the various features the products are ranked separately. Three possible use cases are also executed that covers all possible scenarios- [1] The user searches for a Single Feature. ex- Need a mobile which has awesome battery. The products are ranked based on feature that is present in the user query based on the sentiment computation. [2] The user searches for Multiple Feature then a combination feature is computed and products are ranked. ex- Need a mobile which has awesome battery and nice camera

[3] The user searches query with no feature. ex- Need a good mobile. For this query the total feature vector is computed by using the token based frequency and adding the values. One more innovation along with feature based sentiment computation is that the demography information like City and State based reviews are collected and then recommendations are given based on demographics

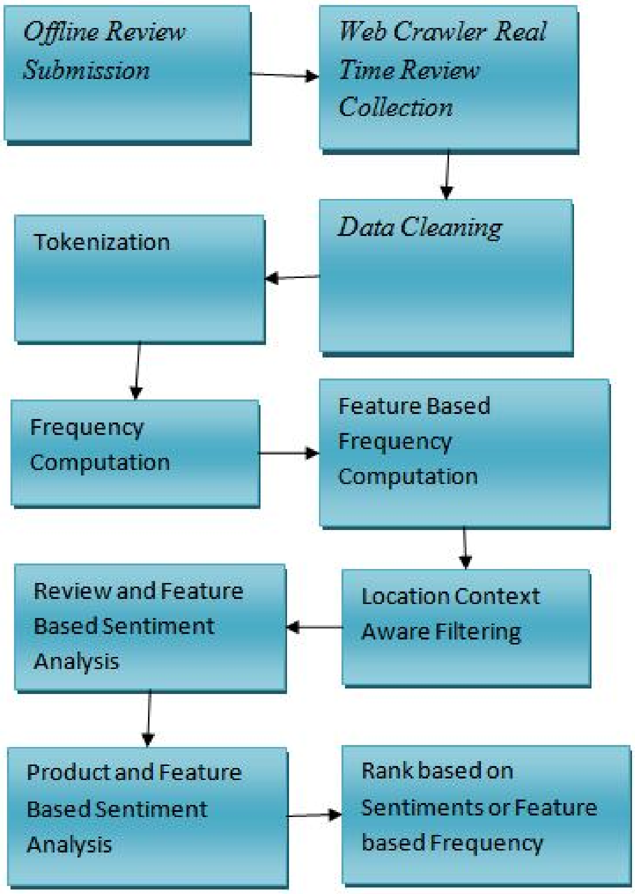


Fig: Proposed System Stages

III.A Offline Review Submission

The Offline Review Submission is a process where a customer gets authenticated and then product name and review is given and then reviews are stored in the format of the following

#	Name
1	**REVIEWID**
2	REVIEWDETAILS
3	PRODUCTID
4	CATID
5	STATE
6	CITY
7	COUNTRY

Review ID- Unique ID for the review
Review Details – The details of the review for the product
Product ID- Unique ID for the product
CATID – Product type
STATE – The State to which user belongs
CITY –Unique City
COUNTRY - The Country to which the user belongs

III.B Online Review Submission

The real time reviews are collected by providing Web URL, Selecting the Product and then providing an Xpath. The reviews are collected from the web page based on the product by applying the DOM based tree creation and extracting the content on the specific DOM nodes where reviews are found. Each web site for collecting reviews from online have their own way of rendering the reviews. Hence Xpath must be configurable. The algorithm can be described as in Algorithm1 Snippet.

Algorithm1

Input

URL, P_i*, Xpath*

Details

1) *Hit the website using the URL*

2) *Download the HTML using Jsoup Parser*

3) *Convert the HTML into Document Object Model (DOM)*

4)

a) *Let* f *be the root of DOM.*

b) *The set* $\{C1, C2, ...Cn\}$ *are childs of* f.

c) *for each of nested childs the traverse is done and the following condition is looked for*

$NC == Xpath$

Where,

NC = Nested Child

d) *if the condition is satisfied then text is extracted for the satified nested child element e*

e.text()

where e is element which has the same xpath

Fig: Algorithm1 Snippet

III.C Data Cleaning

The Data Cleaning algorithm is responsible for removal of stop words. Each of review is cleaned by removing the stop words from reviews. These are the set of words which do not have any specific meaning. The data mining forum has defined set of keywords which do not have any meaning like *a, able, about, across, after, all, almost, also, am, among, an etc*

#	Name
1	CLEANID
2	REVIEWID
3	CLEANREVIEW
4	PRODUCTID
5	PRODUCTTYPE

CLEANID – Unique Id for the clean review
Review ID- Unique ID for the review
Clean Review – Review Description after the review has been cleaned
Product ID- Unique ID for the product
CATID – Product type

The data cleaning uses a set of delimiters like comma, semicolon etc along with set of stop words are used for data cleaning

The algorithm can be described as Algorithm Snippet2

Data Cleaning Algorithm

Data Cleaning Algorithm

Input :

Set of reviews $\{R1, R2, \ldots\ldots\ldots\ldots, Rn\}$

Where

R_i *contains a set* $\{reviewId, reviewDesc, product, city, country, state\}$

Output :

Set of Clean reviews

$\{CR1, CR2, \ldots\ldots\ldots\ldots, CRn\}$

Details :

1) *measure the count of set of reviews* $N_{reviews}$

2) *for each* R_i *in* $\{R1.R2 \ldots\ldots\ldots, Rn\}$

a) obtain the description of Ri known as Si

b) remove all unwanted symbols if present in Si

c) convert S_i *in the form of Queue* Q_i

d) measure the count of number of elements of Q_i *Nq*

e) for each ti in $\{t1, t2, \ldots\ldots tNq\}$

1) *check* $ti \in \{sw1, sw2, \ldots\ldots swn\}$ *if yes then move on to next element of Q*

2) *check* $ti \in \{sw1, sw2, \ldots\ldots swn\}$ *if No then add t to CR move on to next element of Q*

Where ,

t = word

sw = stopword

3) *peform above 2 steps for all tokens and obtain clean review CRi*

3) *After the above steps are done a set of clean reviews are obtained*

$\{CR1, CR2, \ldots\ldots\ldots\ldots, CRn\}$

Fig: Algorithm Snippet2

III.C Tokenization

Tokenization is a process of converting the clean data into a set of words known as tokens. Each of the token can be represented as below

#	Name
1	TOKENID
2	TOKENNAME
3	REVIEWID
4	PRODUCTID
5	PRODUCTTYPE

TOKENID- Unique ID for each word
TOKENNAME- Name of the token
REVIEWID- Unique ID for the review
PRODUCTID-Unique ID for the product
PRODUCTTYPE- Unique ID for the product

Tokenization Algorithm

Input :

Set of clean reviews

$\{CR1, CR2, \ldots\ldots\ldots CRn\}$

Output :

Set of tokens

$\{t1, t2, \ldots\ldots\ldots tn\}$

Details :

1) *Measure the count of clean reviews* CR_{count}

2) *for* $k = 1 : CR_{count}$

a) obtain the kth clean review

b) Convert review CR_k *in FIFO queeue*

c) measure the number of elements of FIFO queue N_{FIFO}

d) for $j = 1 : N_{FIFO}$

1) *obtain the jth Q element* Q_j *and form a set tj*

3) *The set of tokens* $\{t1, t2, \ldots\ldots, tn\}$ *are the final set*

III.D Frequency Computation

This is a process in which the frequency computation is performed. For each of the reviews the frequency is computed. Frequency is number of times a i^{th} token appears in j^{th}. Review The frequency matrix is computed in the following format

#	Name
1	FREQID
2	REVIEWID
3	TOKENNAME
4	FREQ
5	PRODUCTID
6	PRODUCTTYPE

FREQID- Unique ID for each word
TOKENNAME- Name of the token
REVIEWID- Unique ID for the review
FREQ- Count of the word
PRODUCTID-Unique ID for the product
PRODUCTTYPE- Unique ID for the product

The frequency computation can be done as follows

Frequency Computation Algorithm

Input : set of tokens $\{t1, t2, \ldots, tn\}$

Output : $\{w1, w2, \ldots, wm\}$

$m <= n$

m = number of words after freq computation

n = number of words before freq computation

Details :

1) measure the count of set of tokens N_{token}

2) find the unique set of tokens $\{u1, u2, \ldots, uk\}$

$k < n$ if tokens repeat

$k <= n$ if tokens are not repeated

3) measure the count of unique token ui in the set of $\{t1, t2, \ldots, tn\}$ call it as ci

4) Now a map is created with key as u and value as c

III.E Feature Based Computation

The feature based frequency can be computed using a set of features common to all products. It will not maintain the count of other worlds apart from feature words. The feature based frequency can be described as follows

Feature Based Frequency Algorithm

Input :

$\{w1, w2, \ldots, wn\}$

set of words which belong to frequency set

each w has the following

wid : unique id for word w

f : frequency of word w

p : product id

pt : product type

r : review

Output :

$\{o1, o2, \ldots, ot\}$

o are set t of words

t total words

each $o \in \{fw1, fw2, \ldots, fwn\}$

fw is a feature word

Details

1) Find the set of reviews $\{r1, r2, \ldots, rn\}$

2) for each review r

a) find the set of tokens $\{t1, t2, \ldots, tk\}$

b) measure the count of tokens N_{token}

c) for $k = 1 : N_{token}$

1. obtain the kth token tk

2. check $tk \in \{fw1, fw2, \ldots, fwn\}$

if yes

measure count of tk

if no skip

3) finally obtain the value oi

o has the following

oid : id of word

w : feature word

f : frequency

pid : product id

pt : product type

III.F Sentiment Analysis Review Based

In this module for each of the feature and review the sentiments are computed by dividing the entire review into a set of sentences. After that each sentence is checked for feature and if feature is present then negative and positive sentiment of the sentence is determined and likewise process is repeated for all sentences of the review. In the same way the process is repeated for all reviews across the products and the matrix is filled with the following

#	Name
1	SENTID
2	REVIEWID
3	POSITIVERATING
4	NEGATIVERATING
5	NEUTRALRATING
6	PRODUCTID
7	PRODUCTYPE
8	FEATURETYPE

SENTID- Unique ID for row computation
REVIEWID- Unique ID for the review
POSITIVERATING- The positive sentiments for the review per feature type
NEGATIVERATING- The negative sentiments for the review per feature type
PRODUCTID- Unique ID for the product
PRODUCTTYPE- Type of Product
FEATURETYPE- The feature type are like Battery, Camera etc

The set of positive and negative keywords are used to find the sentiments of the sentence. This computation happens only if the sentence has the feature otherwise the sentence is given only neutral sentiment.

The review based sentiment computation algorithm can be described as follows

Review Based Sentiment Analysis

Input: Set of Reviews collected offline or online

$\{R1, R2, \ldots, Rn\}$

Where ,

$Ri = i^{th}$ review

n = number of reviews

Output:

Review based sentiment matrix which is equal to number of reviews

$\{SR1, SR2, \ldots, SRfn\}$

Where .,

$fn = n * f$

f = number of features

$SRi = i^{th}$ sentiment review

Each SR has the following

rid : Rid of Sentiment review

SR : Sentiment review

PS : positive sentiment

NS : negative sentiment

NuS : Nuetral Sentiment

ft : feature type

P : product

Details

1) Count the number of reviews Nc
2) for $k:1--->Nc$
   a) Obtain the k^{th} review

Divide the review into a set of
c) if the sentences can be treated as

$S_f ---> \{s1, s2, ..., sm\}$

$si = i^{th}$ sentence

m = number of sentences

d) for $j--->1:m$

1) Obtain the j^{th} sentence

2) Check whether the sentence sj has ft

$if \ sj \in ft$

PS : Number of positive sentiments

NS : Number of negative sentiments

$if \ PS = NS = 0 \,\|\, sj \notin ft$

Nus : Neutral Sentiments exist

3) Repeat the above process for all sentences

$$PS_{Ri} = \sum_{j=1}^{m} PS_j$$

$$NS_{Ri} = \sum_{j=1}^{m} NS_j$$

$$NuS_{Ri} = \sum_{j=1}^{m} NuS_j$$

Where,

PS_{Ri} = positive sentiments for i^{th} review

NS_{Ri} = negative sentiments for i^{th} review

NuS_{Ri} = nuetral sentiments for i^{th} review

PS_j = positive sentiment for j^{th} sentence

NS_j = negative sentiments for j^{th} sentence

NuS_j = nuetral sentiments for j^{th} sentence

d) The sentiment set is computed as follows

$SR_i = \{rid, SR, PS, NS, NuS, P\}$

III.G Product based Sentiment Reviews

The Product based sentiment computation is performed using the following

Find the unique set of products for which reviews are collected

The following is computed for each product

Find the reviews for the product. For each of the review per feature the following is computed

The positive sentiment for the product is computed using the following

$$PS_{product, f_j} = \sum_{i=1}^{N_{reviews}} PS_i$$

Where ,

$N_{reviews}$ = Number of Re views

PS_i = Positive Sentiment for i^{th} review

$f_j = j^{th}$ feature

$0 \le f \le N_f - 1$

N_f = Number of features

The negative sentiment for the product is computed using the follow

$$NS_{product, f_j} = \sum_{i=1}^{N_{reviews}} NS_i$$

Where ,

$N_{reviews}$ = Number of Re views

NS_i = Negative Sentiment for the i^{th} review

$f_j = j^{th}$ feature

$0 \le f \le N_f - 1$

N_f = Number of features

The neutral sentiment is computed using the following

$$NU_{product, f_j} = \sum_{i=1}^{N_{reviews}} NU_i$$

Where ,

NU_i = Neutral Sentiments for i^{th} review

$N_{reviews}$ = Number of reviews

$f_j = j^{th}$ feature

$0 \le j \le N_f$

N_f = Number of features

The algorithm can be defined as follows

Product Based Sentiments

Input: Review based sentiment matrix

$\{SR_1, SR_2, ..., SR_{fn}\}$

Where ,

$SR_i = i^{th}$ sentiment review

fn = Number of rows of sentiment matrix

Each SR has the following

rid : unique identifier for row

R_i : review id

P : product

PS : positive sentiments

NS : negative sentiments

NuS : nuetral sentiments

Output: Product based Sentiments

$\{PS_1, PS_2, ..., PSn\}$

Where ,

$PS_i = i^{th}$ product sentiment

n = number of products

Each PS has the following properties

P : product

PS : positive sentiments

NS : negative sentiments

NuS : neutral sentiments

ft : feature type

Details:

1) Find the number of unique products N_u
2) for $k:1--->N_u$
   a) obtain the k^{th} product
   b) Find the list of reviews $\{R_1, R_2, ..., R_m\}$

c) obtain the total positive sentiments

$$PS_{product,f_j} = \sum_{i=1}^{N_{reviews}} PS_i$$

Where,

$N_{reviews} = Number\ of\ Reviews$

$PS_i = Positive\ Sentiment\ for\ i^{th}\ review$

$f_j = j^{th}\ feature$

$0 \le f \le N_f - 1$

$N_f = Number\ of\ features$

d) Obtain the negative sentiments

$$NS_{product,f_j} = \sum_{i=1}^{N_{reviews}} NS_i$$

Where,

$N_{reviews} = Number\ of\ Reviews$

$NS_i = Negative\ Sentiment\ for\ the\ i^{th}\ review$

$f_j = j^{th}\ feature$

$0 \le f \le N_f - 1$

$N_f = Number\ of\ features$

e) obtain the neutral sentiments

$$NU_{product,f_j} = \sum_{i=1}^{N_{reviews}} NU_i$$

Where,

$NU_i = Neutral\ Sentiments\ for\ i^{th}\ review$

$N_{reviews} = Number\ of\ reviews$

$f_j = j^{th}\ feature$

$0 \le j \le N_f$

$N_f = Number\ of\ features$

III.H Rank Based on Sentiments per Feature

The user is allowed to enter a query and only the review sentiments computed for the specific region are taken into consideration and then finally the products are ranked based on having positive sentiment maximum, negative sentiment minimum and neutral sentiment maximum.

If it contains more than one feature then all the features negative sentiments, positive sentiments and neutral sentiments per product are added and finally ranking s performed.

The ranking algorithm can be described as follows as

Ranking based on Sentiments

Input: Query for the User Q_u

Output: Ranked Products based on Sentiments

$\{P_1, P_2, ..., P_n\}$

P: product

Order is based on Positive Sentiments Maximum, Negative Minimum and Neutral Sentiments maximum

Details:

1) Check the user entered query Q contains the feature

2) find the features $\{f1, f2, ... fq\} \in afs$

Where, f_i is the i^{th} feature and

$afs\ is\ actual\ feature\ set\ of\ product$

3) find the number of features present in search query N_{fq}

4) if $N_{fq} = 1$ then find the product set recommendations based on single feature

$\{ps1, ps2, ...psn\}$

This has maximum positive, minimum negative, maximum neutral and maximum feature. psi has the ranked product

5) if $N_{fq} > 1$ then find the set of unique product s from sentiment matrix and then perform the additive measure across multiple features

For each $ptemp ---> 1 : N_p$

$$T_{ptemp,positive} = \sum_{i=1}^{N_{fq}} PS(i)$$

$$T_{ptemp,negative} = \sum_{i=1}^{N_{fq}} NS(i)$$

$$T_{ptemp,neutral} = \sum_{i=1}^{N_{fq}} NuS(i)$$

Find the sentiment index

$SI(ptemp) ---> \{p, Tp, Tn, Tnu\}$

Sort the products based on sentiment index and recommend products

III.I Ranking Algorithm based on No Feature

If the query does not have any feature then the entire query is divided into tokens and then frequency of those tokens across products are found and then they are finally added together. The algorithms are described as follows

Ranking based on No Feature

Input: Query Q

Output: Product ranking set

$\{pf1, pf2, ... pfn\}$

$Where,\ pfi = i^{th}\ product\ ranked\ based\ on\ feature$ Details

1) Divide the searched query into a set of tokens $\{t1, t2, ..., ts\}$

2) Find the set of unique products which have to be ranked

$\{p1, p2, p3, ..., Pnq\}$

3) for each of product pk the reviews are found and then the matrix of correlation is found as follows per product

Let $\{r1, r2, ..., rn\}$ are set of reviews associated with product pk then a matrix can be constructed which depicts the correlation

P1	t1	ts-1	ts	TC
r1	F11	F1s-1	F1s	Tc1
r2				Tc2
rn	Fn1	Fn2	fns	Tcn

Where
fij = Number of times appears in jth review
Tci= Total correlation for ith review
4) After matrix are computed for all the products the total correlation per product is computed using

$$TC_p = \sum_{i=1}^{n} TCi$$

Where ,

n = *number of reviews*

TC_p = *total correlation of product*

TC_i = *total correlation per review for p*

5) All the products are arranged in descending order of total correlation and recommended for user

IV. Experiment Results and Analysis

In this work 2 web applications are created one for implementation of existing approaches namely content based filtering, collaborative based filtering, Pearson recommendations and sentiment based recommendations. The second application is responsible for proposed approach implementation which involves demographic based review collection from consumers as well as from N websites like Amazon, flipkart , snap deal etc. After that each processing steps of the algorithms as described in the previous section have been implemented.

Collaborative Based Recommendations

In this type of recommendations whether it is registered user or unregistered user numeric rating is collected across N number of users. For experiment purpose 50 mobiles are considered with 5 million training users on board.

Rating Submission by Users

The following screen is used in order to provide the ratings for the product by the user. Any user selects the product and rates the product on the scale of 1-5.

Fig1: Rating Submission

Fig1 shows the rating submission screen where product named SAMSUNG GALAXY S1 is selected and rating of 5 is given and submitted. In the same way N number of users submits their respective ratings.

Collaborative Rating of Users

Ratings Output

Product ID	Product NAME	RATING
3	NOKIA LUMINA	13000
4	LG	9000
2	SAMSUNG GALAXY S3	8000
5	Apple IPhone 6	5567
1	SAMSUNG GALAXY S1	4000

Fig: Collaborative Rating

The above grid shows the total aggregated rating across the users for all the products and the products are ranked based on the aggregated rating. Note aggregated rating is the sum of all the ratings given by the users per product. As one can see from the fig Nokia Lumina is the product suggested to the user followed by LG, Samsung Galaxy S3, Apple IPhone 6 and Samsung Galaxy S1 which is ordered based on the aggregated rating. Nokia Lumina has the highest rating of 13000

Collaborative Rating Variation

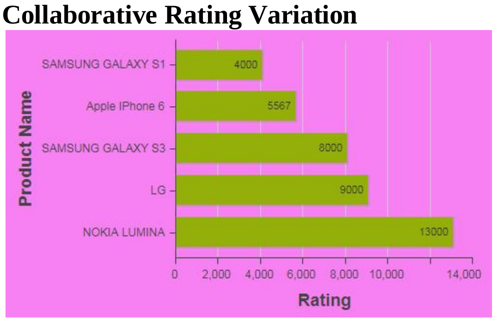


Fig: Collaborative Rating Graph

Fig shows the collaborative rating graph where more horizontal indicates good product. Because Nokia Lumina has high aggregated rating the horizontal of Nokia Lumina is more followed by LG which has next aggregated rating of 9000.

Content Based Recommendations

The content based recommendations are provided only for registered users and also who have a valid session. For content based recommendations history profile is created based on product purchases /transaction history for the user.

Personal Rating Module

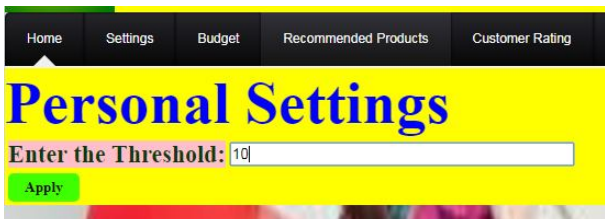


Fig: Personal Settings

Fig shows the personal settings which is set by user as a threshold to filter the products. As shown in the fig user has set a value has 10.

Content Based Recommendations

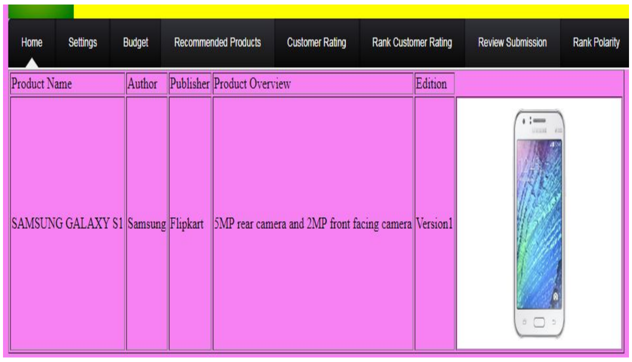


Fig: Content Based Recommendations

Fig shows Content Based recommendations provided to the user. As shown in fig Samsung Galaxy S3 is proposed because the user has done at least 10 or more transactions on the same product. During each transaction of the user the merchant stores 2 important factors order info and order details.

The order info has the following columns ORDERID, LOGINID, ORDERDATE, TOTALAMOUNT and EMAIL.

ORDERID	LOGINID	ORDERDATE	TOTALAMOUNT	EMAIL
1	SACHIN1245	2016-03-22	5000	sachin@gmail.com
2	SACHIN1245	2016-03-22	5000	sachin@gmail.com
3	sehwag123	2016-04-03	5000	sehwag@gmail.com
4	sehwag123	2016-04-03	5000	sehwag@gmail.com
5	sehwag123	2016-04-03	1500	sehwag@gmail.com
6	sharukh1234	2016-04-10	5000	sharukh@gmail.com

Order ID is unique generated number for each of the transaction, Login ID is the login credentials of the user, Order Date is the date of purchase order and Total Amount is the amount spends by the customer for the purchase of product. Email is the email of the user provided during the registration process

The Order Details contains the following

ORDERID	PRODUCTID	QUANTITY
1	1	1
2	1	1
3	1	1
4	1	1
5	3	1
6	1	1
7	1	1
8	1	1

Order ID is the unique transaction id, product id is the unique id for the product and quantity is the count of purchases of the same product within a transaction.

Pearson Correlation

In Pearson correlation the rating of the product is collected for a given product similar to collaborative recommendations but only from registered users. The data is maintained in the following matrix format in order to devise the recommendations from the Pearson formula. After the rating is submitted by registered users it is stored in matrix format

USERNAME	RATING	PRODUCTID
SACHIN1245	3	1
SACHIN1245	5	2
SACHIN1245	3	3
SACHIN1245	5	5
SACHIN1245	1	2
aaquibnawazrvce	5	5
aaquibnawazrvce	4	1
aaquibnawazrvce	5	2
aaquibnawazrvce	4	3
aaquibnawazrvce	2	4
SACHIN1245	1	3
sehwag123	4	1
sehwag123	4	3
sehwag123	4	4
dhoni123	3	1

Username is the name of the registered user who provided the rating. Rating is the numeric value given by user and finally product id is the unique id associated with the product. These values are substitute in pearson correlation and the predicted rating is computed and products are ranked

Product Name	Product ID	Predicted Ratings
NOKIA LUMINA	3	0.933333333333333
Apple I Phone6	5	0.166666666666667
SAMSUNG GALAXY S1	1	-0.233333333333333
SAMSUNG GALAXY S3	2	-0.666666666666667
LG	4	-1.4

Fig: Pearson Recommendations

Fig shows Pearson recommendations as shown in fig Nokia Lumina is recommended to the user as it has highest predicted rating followed by Apple I Phone6 and then remaining products are ranked.

PROPOSED METHOD ANALYSIS

Review Submission Customers

The review submission is done by selecting the product and then review description is added and submitted by the user.

Fig: Review Submission

Fig shows the user has selected Samsung Galaxy S1 and user has entered the review details as “Samsung Galaxy S1 is an awesome battery mobile.". Like this many users can provide the reviews across the products

Online Review Collection using Web Crawler

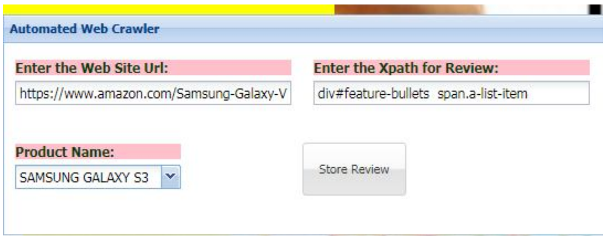


Fig: Online Review Collection

Fig shows online review collection process. The user enters web site URL of amazon where reviews are present, Xpath the HTML depth of reviews and Selects the product name as Samsung Galaxy S3.One the parameters are submitted Web Crawler algorithm is triggered which converts the HTML into a DOM tree. After that the DOM is navigated and reviews are extracted dynamically and then a set is formed in the format {ReviewId, ProductId, ProductName, Review Description}

Review Collection Matrix

The review collection matrix can be described as below

Reviews Collected using Web Crawler

Review ID	Product Name	Review Details	City	State	Country
1	SAMSUNG GALAXY S1	Samsung galaxy s1 is an awesome mobile it has nice camera and super battery.Samsun...	Newyork	California	United States
2	NOKIA LUMINA	Nokia Lumina has an awesome touch it has awesome battery. It has good memory and i...	Newyork	California	United States
3	SAMSUNG GALAXY S1	Hi Samsung Galaxy S1 is an awesome mobile. It has nice camera and good memory. It h...	Dehli	Dehli	India
4	NOKIA LUMINA	Nokia Lumina is a nice mobile. It has awesome sound quality. It has nice touch. It has g...	Dehli	Dehli	India
5	SAMSUNG GALAXY S1	Okay, I'm coming from a Motorola Cliq XT running the ancient and completely obsolete ...	Mumbai	Maharastra	India
6	SAMSUNG GALAXY S1	I was an iPhone user who decided to buy new SGS4G for some "freedom" and cheaper c...	Mumbai	Maharastra	India
7	NOKIA LUMINA	Seriously, you cannot buy a better smartphone at this price. I purchased mine at my loc...	Bangalore	Karnataka	India
8	NOKIA LUMINA	Was not happy with the product. I returned if after the first use. The processor is slow, t...	Bangalore	Karnataka	India
9	SAMSUNG GALAXY S1	Samsung Galaxy s1 has awesome battery backup. It has nice battery backup.Samsung G...	Bangalore	Karnataka	India
10	NOKIA LUMINA	Nokia Lumina has a bad battery. It has really very nice memory. It has very bad battery....	Bangalore	Karnataka	India
11	LG	LG is a nice mobile. It has nice memory. It has bad battery.LG is a nice mobile. It has nic...	Bangalore	Karnataka	India
12	SAMSUNG GALAXY S1	Samsug Glalaxy S1 is a nice mobile. It has awesome amount of storage. It has very nice ...	Bangalore	Karnataka	India
13	SAMSUNG GALAXY S1	This is test mobile. It has nice battery life it has awesome battery life. It has very good b...	Mumbai	Delhi	Afghanistan

Fig: Reviews collected using Web Crawler

Fig shows the review collected using Web Crawler algorithm Review ID is the unique id for the review, Product Name of the product for which review is collected, Review Details of the reviews collected either by review submission or by online review collection, City, State and Country are the new elements.

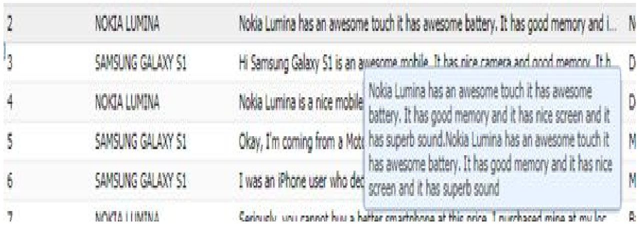
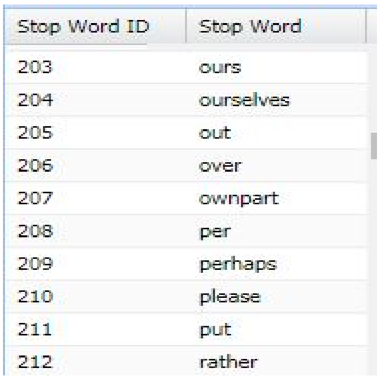


Fig: Detailed Review View

Fig shows the detailed review view. When the user hovers on the review full review description is made

Stop Word Analysis

The stop words are list of words given by the web mining forums and can be viewed in application as below

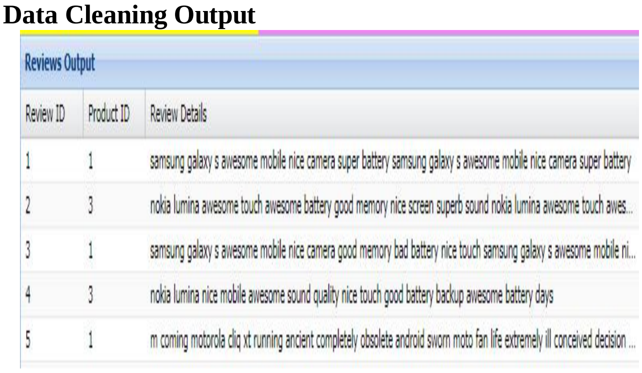


Stop Word ID	Stop Word
203	ours
204	ourselves
205	out
206	over
207	ownpart
208	per
209	perhaps
210	please
211	put
212	rather

Fig: Stop Word List

Fig shows the list of stop words.

Data Cleaning Output



Reviews Output

Review ID	Product ID	Review Details
1	1	samsung galaxy s awesome mobile nice camera super battery samsung galaxy s awesome mobile nice camera super battery
2	3	nokia lumina awesome touch awesome battery good memory nice screen superb sound nokia lumina awesome touch awes...
3	1	samsung galaxy s awesome mobile nice camera good memory bad battery nice touch samsung galaxy s awesome mobile ni...
4	3	nokia lumina nice mobile awesome sound quality nice touch good battery backup awesome battery days
5	1	m coming motorola cliq xt running ancient completely obsolete android sworn moto fan life extremely ill conceived decision ...

Fig: Clean Reviews Output

Fig shows the data cleaning output. As shown in the fig there are 3 values namely review id, product id and review details. The review detail statements do not contain even a single stop word.

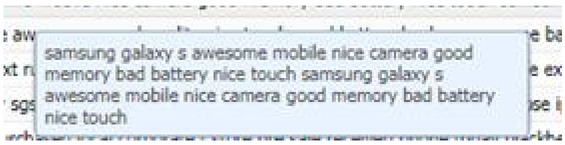


Fig: Detail Review Output

Fig shows the detail review output. When user hovers on the review it gives detailed description.

Tokenization

Tokenization output contains list of values as follows

View Tokens Output

Review ID	Product ID	Token ID	Token Name
1	1	3731	samsung
1	1	3732	galaxy
1	1	3733	s
1	1	3734	awesome
1	1	3735	mobile
1	1	3736	nice

Fig: Tokenization Output

Fig ajsda Fig shows the Tokenization Output as shown in the fig there are 4 values namely Review ID, Product ID, Token ID and Token Name. Review ID is the unique ID for the review. Product ID is the id of the product, Token ID is the unique ID for the word. Token Name is the word.

Frequency Output

View Frequency Output

Review ID	Product ID	Token Name	Frequency ID	Frequency
1	1	camera	2378	2
1	1	super	2379	2
1	1	battery	2380	2
2	3	nokia	2381	2
2	3	lumina	2382	2
2	3	awesome	2383	4

Fig: Frequency Output

Fig shows the Frequency Output. As shown in the fig there are values namely Review ID, Product ID, Token Name, Frequency ID and Frequency. Review ID is the unique ID for the review. Product ID is the unique ID for the product. Token Name is the word, Frequency ID is the unique auto generated value for each word and Frequency is the count of number of times word is repeated.

Feature Vector Computation

This module is responsible for computing feature based frequency based on a set of features namely Battery, Camera, Memory, Touch, Sound and Screen.

Feature Output

Review ID	Product ID	Feature Type	Feature Based Freq
2	3	MEMORY	2
2	3	BATTERY	2
2	3	TOUCH	2
2	3	SOUND	2
2	3	SCREEN	2
4	3	CAMERA	0
4	3	MEMORY	0
4	3	BATTERY	2
4	3	TOUCH	1
4	3	SOUND	1

Fig: Feature Vector Matrix

Fig shows the sub part of feature matrix. Each row is a set of {ReviewId, ProductID, Feature .Type, Feature Based Freq}. The first row has the value {2,3,MEMORY,2}. The value 2 represents the Review ID, 3 is the Product ID, MEMORY is the feature type and Feature Based Freq has the value 2. Like this the computation is made across all products and feature types for all reviews.

Review Biased Sentiment Analysis

The review based sentiment analysis can be described as follows

Polarity Computation Output

Review ID	Product ID	Positive Rating	Negative Rating	Neutral Rating	Feature Type
1	1	2	0	0	CAMERA
1	1	2	0	0	MEMORY
1	1	4	0	0	BATTERY
1	1	4	0	0	TOUCH
1	1	4	0	0	SOUND
1	1	4	0	0	SCREEN
2	3	0	0	1	CAMERA
2	3	4	0	1	MEMORY
2	3	6	0	1	BATTERY
2	3	8	0	1	TOUCH

Fig: Review based Polarity Matrix

Fig shows Review Based Polarity Matrix. As shown in the fig the polarity matrix has the following entities namely Review ID, Product ID, Positive Rating, Negative Rating, Neutral Rating and Feature Type. The 1st row has the following {1,1,2,0,0,CAMERA} has the value of Review ID as 1 ,Product ID has the value 1, Positive Rating has the value of 2, Negative Rating has the value of 0 and Feature Type as CAMERA.

Product Based Sentiment Matrix

The Product Based Sentiment Matrix can be described as follows

Total Polarity Output

Product ID	Positive Rating	Negative Rating	Neutral Rating	Feature Type	Total Feature
1	4	0	5	CAMERA	8
1	6	0	10	MEMORY	4
1	22	2	12	BATTERY	29
1	23	2	13	TOUCH	2
1	23	2	14	SOUND	0
1	24	2	14	SCREEN	6
3	0	0	5	CAMERA	0
3	4	0	9	MEMORY	4
3	8	6	11	BATTERY	14
3	10	6	13	TOUCH	3

Fig: Polarity Computation

Fig shows the product based polarity matrix which will contain {ProductID, Positive Rating, Negative Rating, Neutral Rating and Feature Type}. Product ID is the ID for Product, Positive Rating is the how much total positive polarity the product is having by adding positive rating across all reviews, Negative Rating is the total negative polarity the product is having and Neutral rating is the total neutral rating of the product. Feature Type is the name of the feature. Total Feature is the number of times feature word is repeated across all reviews for the product .The first row has the value 1 for product id, 4 is the positive rating, 0 is the negative rating, 5 is the neutral rating, feature type has the value CAMERA and Total Feature is having the value of 8.

Product Graphs

The following table contains the following values

ProductInfo	
Product ID	Product Name
1	SAMSUNG GALAXY S1
2	SAMSUNG GALAXY 3
3	NOKIA LUMINA
4	LG
5	NOKIA 6350
6	One Plus One
7	Appo Phone

Fig: Product Information List

Fig shows the product information list which has the values Product ID and Product Name. The 1st row has the following 1 as Product ID and Product Name as Samsung Galaxy S1.

Battery Polarity Graphs

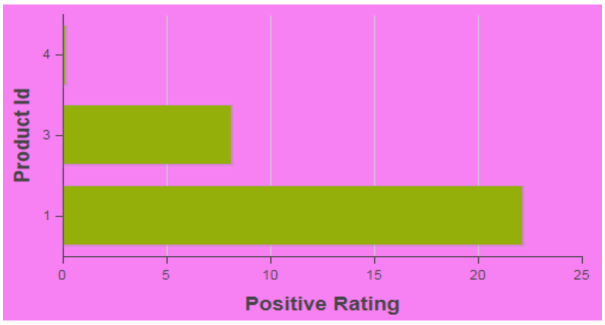


Fig: Positive Rating Battery Graph

Fig shows the positive rating battery graph which has x axis as the value of polarity and y axis as the value of product ids. For Sentiment analysis graph a set of 4 products reviews are collected and analyzed there polarity for battery. Product which has the ID as 1 has a value of 22 , Product 3 has the value of 8 and Product 4 has the value of 0 as Positive Rating. No reviews are collected for 2nd product hence it is not seen in the graph.

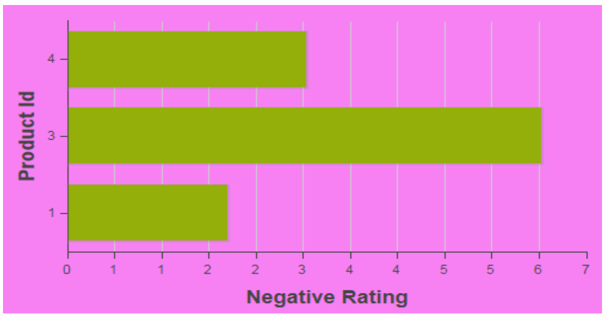


Fig: Negative Rating Battery Graph

Fig shows the negative rating battery graph which has x axis as the value of polarity and y axis as the value of product ids. For Sentiment analysis graph a set of 4 products reviews are collected and analyzed there polarity for battery. Product which has the ID as 1 has a value of 2, Product 3 has the value of 6 and Product 4 has the value of 3 as Negative Rating. No reviews are collected for 2nd product hence it is not seen in the graph.

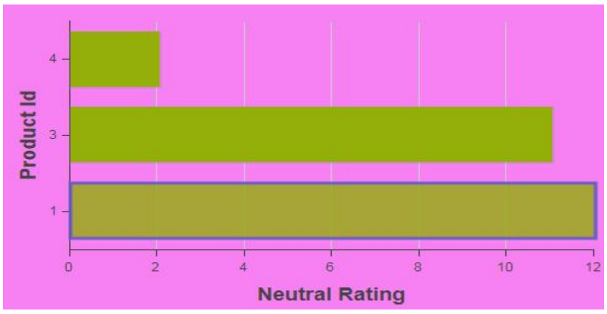


Fig: Neutral Rating Battery Graph

Fig shows the Neutral rating battery graph which has x axis as the value of polarity and y axis as the value of product ids. For Sentiment analysis graph a set of 4 products reviews are collected and analyzed there polarity for battery. Product which has the ID as 1 has a value of 12, Product 3 has the value of 11 and Product 4 has the value of 2 as Neutral Rating. No reviews are collected for 2nd product hence it is not seen in the graph.

Note- In a similar fashion for all the remaining features graphs are computed namely Camera, Memory, Touch, Sound and Screen.

Performance Analysis

Precision Measure

The Precision Measure is defined as below

$$P = \frac{TP}{TP + TN}$$

Where ,

P = *precision*

TP = *Total Positive*

TN = *Total Negative*

Recall Measure

Recall Measure is defined as below

$$RM = \frac{TP}{TP + FN}$$

Where ,

RM = Re *call Measure*

FN = *False Negative*

FI Measure

FI Measure can be defined as follows

$$FI = \frac{2 * RM * P}{RM + P}$$

In order to do performance measure reviews across various products are considered

Parameters	Polarity Algorithm [48]	Feature based Polarity Algorithm
Precision	25.93%	28.85%
Recall	1.49%	4.5%
FI Measure	2.81%	7.78%

REFERENCES

[1] A.G. Aliyev ; V.A. Abbasova ; E.H. Musaeva,"Analysis of the basic software for the electronic commerce system",Application of Information and Communication Technologies (AICT), 2011 5th International Conference

[2] Peng Yu, Collaborative filtering recommendation algorithm based on both user and item,Computer Science and Network Technology (ICCSNT), 2015 4th International Conference

[3] Hui Li ; Fei Cai ; Zhifang Liao, Content-Based Filtering Recommendation Algorithm Using HMM, and Information Sciences (ICCIS), 2012 Fourth International Conference ,17-19 Aug. 2012

[4] Leily Sheugh ; Sasan H. Alizadeh, A note on pearson correlation coefficient as a metric of similarity in recommender system,AI & Robotics (IRANOPEN), 2015

[5] K. Mouthami ; K. Nirmala Devi ; V. Murali Bhaskaran,Sentiment analysis and classification based on textual reviews,Information Communication and Embedded Systems (ICICES), 2013 International Conference,29th april

[6] Sun Lin ; Xu Wenzheng,E-Commerce Personalized Recommendation System Based on Web Mining Technology Design and Implementation, Intelligent Transportation, Big Data and Smart City (ICITBS), 2015 International Conference, 21 January 2016

[7] Sanjeevan Sivapalan ; Alireza Sadeghian ; Hossein Rahnama ; Asad M. Madni,"Recommender systems in e-commerce",World Automation Congress (WAC), 2014

[8] Ruisheng Zhang ; Qi-dong Liu ; Chun Gui ; Jia-Xuan Wei ; Huiyi Ma,"Collaborative Filtering for Recommender Systems",Advanced Cloud and Big Data (CBD), 2014 Second International Conference,20-22 Nov. 2014

[9] Feng Zhao ; Fengwei Yan ; Hai Jin ; Laurence T. Yang ; Chen Yu,"Personalized Mobile Searching Approach Based on Combining Content-Based Filtering and Collaborative Filtering",IEEE Systems Journal , Volume: 11, Issue: 1, March 2017

[10] Natedao Thotharat,"Thai local product recommendation using ontological content based filtering",Knowledge and Smart Technology (KST), 2017 9th International Conference, 1-4 Feb. 2017

[11] Praveena Mathew ; Bincy Kuriakose ; Vinayak Hegde,"Book Recommendation System through content based and collaborative filtering method",Data Mining and Advanced Computing (SAPIENCE), International Conference,16-18 March 2016

[12] Mohammed Nazim uddin ; Jenu Shrestha ; Geun-Sik Jo,"Enhanced Content-Based Filtering Using Diverse Collaborative Prediction for Movie Recommendation",Intelligent Information and Database Systems, 2009. ACIIDS 2009. First Asian Conference,1-3 April 2009

[13] Y. Blanco-Fernandez ; J. J. Pazos-Arias ; A. Gil-Solla ; M. Ramos-Cabrer ; M. Lopez-Nores,"Providing Entertainment by Content-based Filtering and Reasoning in Intelligent Recommender Systems",Consumer Electronics, 2008. ICCE 2008. Digest of Technical Papers. International Conference,9-13 Jan. 2008

[14] J. Salter, N. Antonopoulos, "CinemaScreen recommender agent: combining collaborative and content-based filtering",IEEE Intelligent Systems , Volume: 21, Issue: 1, Jan.-Feb. 2006

[15] N. Churcharoenkrung ; Y.S. Kim ; B.H. Kang, "Dynamic Web content filtering based on user's knowledge", Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference,4-6 April 2005

[16] K. Iwahama ; Y. Hijikata ; S. Nishida, "Content-based filtering system for music data",Applications and the Internet Workshops, 2004. SAINT 2004 Workshops. 2004 International Symposium,26-30 Jan. 2004

[17] Monica M Ingle ; M. Emmanues,"Evaluations on sentiment analysis of micro blogging site using topic modeling",Signal Processing, Communication, Power and Embedded System (SCOPES), 2016 International Conference,3-5 Oct. 2016

[18] Venkata Sasank Pagolu ; Kamal Nayan Reddy ; Ganapati Panda ; Babita Majh, "Sentiment analysis of Twitter data for predicting stock market movements",Signal Processing, Communication, Power and Embedded System (SCOPES), 2016 International Conference,3-5 Oct. 2016

[19] Mauro Dragoni ; Giulio Petrucci,"A Neural Word Embeddings Approach For Multi-Domain Sentiment Analysis",IEEE Transactions on Affective Computing,Volume: PP, Issue: 99,21 June 2017

[20] A. M. Abirami ; V. Gayathri,"A survey on sentiment analysis methods and approach",Advanced Computing (ICoAC), 2016 Eighth International Conference,19-21 Jan. 2017

[21] Rupal Bhargava ; Yashvardhan Sharma,"MSATS: Multilingual sentiment analysis via text summarization",Cloud Computing, Data Science & Engineering - Confluence, 2017 7th International Conference,12-13 Jan. 2017

[22] Abhijit Janardan Patankar ; Kshama V. Kulhalli ; Kotrappa Sirbi, "Emotweet: Sentiment Analysis tool for twitter",Advances in Electronics, Communication and Computer Technology (ICAECCT), 2016 IEEE International Conference,2-3 Dec. 2016

[23] Falguni Gupta ; Swati Singal,"Sentiment analysis of the demonitization of economy 2016 India, Regionwise", Cloud Computing, Data Science & Engineering - Confluence, 2017 7th International Conference,12-13 Jan. 2017

[24] Hamidreza Keshavarz ; Mohammad Saniee Abadeh,"MVP: Memetic Voter Patterns for aspect extraction in sentiment analysis", Swarm Intelligence and Evolutionary Computation (CSIEC), 2017 2nd Conference, 7-9 March 2017

[25] Srishti Sharma ; Shampa Chakraverty,"Tracking Context Switches in Text Documents and Its Application to Sentiment Analysis", Micro-Electronics and Telecommunication Engineering (ICMETE), 2016 International Conference,22-23 Sept. 2016

[26] Anshuman ; Shivani Rao ; Misha Kakkar,"A rating approach based on sentiment analysis",Cloud Computing, Data Science & Engineering - Confluence, 2017 7th International Conference,12-13 Jan. 2017

[27] Abdalraouf Hassan ; Ausif Mahmood,"Deep Learning approach for sentiment analysis of short texts",Control, Automation and Robotics (ICCAR), 2017 3rd International Conference,24-26 April 2017

[28] Zamahsyari, Arif Nurwidyantoro, "Sentiment analysis of economic news in Bahasa Indonesia using majority vote classifier", Data and Software Engineering (ICoDSE), 2016 International Conference,26-27 Oct. 2016

[29] Chloé Clavel ;Zoraida Callejas, "Sentiment Analysis: From Opinion Mining to Human-Agent Interaction",IEEE Transactions on Affective Computing (Volume: 7, Issue: 1, Jan.-March 1 2016), Page(s): 74 – 93

[30] Desheng Dash Wu ; Lijuan Zheng ; David L. Olson,"A Decision Support Approach for Online Stock Forum Sentiment Analysis",IEEE Transactions on Systems, Man, and Cybernetics: Systems (Volume: 44, Issue: 8, Aug. 2014)

[31] Bradley Meyer ; Marwan Bikdash ; Xiangfeng Dai, "Fine-grained financial news sentiment analysis",SoutheastCon, 2017,30 March-2 April 2017

[32] Lorenzo Gatti ; Marco Guerini ; Marco Turchi,"SentiWords: Deriving a High Precision and High Coverage Lexicon for Sentiment Analysis",IEEE Transactions on Affective Computing (Volume: 7, Issue: 4, Oct.-Dec. 1 2016),Page(s): 409 – 421

[33] Mukta Y. Raut ; Mayura A. Kulkarni,"Polarity shift in opinion mining",Advances in Electronics, Communication and Computer Technology (ICAECCT), 2016 IEEE International Conference,2-3 Dec. 2016

[34] Deebha Mumtaz ; Bindiya Ahuja,"Sentiment analysis of movie review data using Senti-lexicon algorithm",Applied and Theoretical Computing and Communication Technology (iCATccT), 2016 2nd International Conference,21-23 July 2016

[35] Renato S. C. da Rocha ; Leonardo Forero ; Harold de Mello ; Manoela Kohler ; Marley Vellasco,"Polarity classification on web-based reviews using Support Vector Machine",Computational Intelligence (LA-CCI), 2016 IEEE Latin American Conference,2-4 Nov. 2016

[36] Nipuna Upeka Pannala ; Chamira Priyamanthi Nawarathna ; J. T. K. Jayakody ; Lakmal Rupasinghe ; Kesavan Krishnadeva,"Supervised Learning Based Approach to Aspect Based Sentiment Analysis",Computer and Information Technology (CIT), 2016 IEEE International Conference,8-10 Dec. 2016.

[37] Jyoti Ramteke ; Samarth Shah ; Darshan Godhia ; Aadil Shaikh,"Election result prediction using Twitter sentiment analysis",Inventive Computation Technologies (ICICT), International Conference,26-27 Aug. 2016

[38] Fatih Sağlam ; Hayri Sever ; Burkay Genç,"Developing Turkish sentiment lexicon for sentiment analysis using online news media",Computer Systems and Applications (AICCSA), 2016 IEEE/ACS 13th International Conference,29 Nov.-2 Dec. 2016

[39] Ashraf Elnagar,"Investigation on sentiment analysis for Arabic reviews",Computer Systems and Applications (AICCSA), 2016 IEEE/ACS 13th International Conference, 29 Nov.-2 Dec. 2016

[40] Ritesh Srivastava ; M.P.S. Bhatia,"Ensemble methods for sentiment analysis of on-line micro-texts", Recent Advances and Innovations in Engineering (ICRAIE), 2016 International Conference,23-25 Dec. 2016

[41] Hina Gupta ; Nitasha Hasteer ; Rana Majumdar,"Interpretive structure modelling(ISM) for feature dependency in sentiment analysis", Cloud Computing, Data Science & Engineering - Confluence, 2017 7th International Conference,12-13 Jan. 2017

[42] Wanxiang Che ; Yanyan Zhao ; Honglei Guo ; Zhong Su ; Ting Liu,"Sentence Compression for Aspect-Based Sentiment Analysis",IEEE/ACM Transactions on Audio, Speech, and Language Processing (Volume: 23, Issue: 12, Dec. 2015)

[43] Zhao Jianqiang ; Gui Xiaolin,"Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis"IEEE Access (Volume: 5),Page(s): 2870 - 2879,22 February 2017

[44] Vinaya B. Savadekar ; Manoj E. Patil,"Improved recommendation system with review analysis",Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), 2016 International Conference, 22-24 Dec. 2016.

[45] P Devika ; R C Jisha ; G P Sajeev,"A novel approach for book recommendation systems",Computational Intelligence and Computing Research (ICCIC), 2016 IEEE International Conference,15-17 Dec. 2016

[46] Sundus Ayyaz ; Usman Qamar,"Improving collaborative filtering by selecting an effective user neighborhood for recommender systems",Industrial Technology (ICIT), 2017 IEEE International Conference,22-25 March 2017

[47] Raja Sarath Kumar Boddu,"An integrated assessment approach to different collaborative filtering algorithms",Big Data (Big Data), 2016 IEEE International Conference,5-8 Dec. 2016

[48] Asma Musabah Alkalbani, Ahmed Mohamed Ghamry, Farookh Khadeer Hussain, Omar Khadeer Hussain "Sentiment Analysis and classification for Software as a Service Reviews",2016 IEEE 30th International Conference on Advanced Information Networking and Applications