ISSN No:-2456-2165

Aspect and Review Based Recommendation System

Sujata Khedkar
Associate Professor
Department of Computer Engineering
V.E.S. Institute of Technology
Mumbai-400074

Mihir Joshi
Computer Engineering
V.E.S. Institute of Technology
Mumbai-400074

Mugdha Govilkar
Computer Engineering
V.E.S. Institute of Technology
Mumbai-400074

Tejas Ingale
Computer Engineering
V.E.S. Institute of Technology
Mumbai-400074

Abhishek Kuvar
Computer Engineering
V.E.S. Institute of Technology
Mumbai-400074

Abstract:- The instinctive medium that humans use for communication is of words and not numbers. Ratings and reviews are the most common form of feedback that customers of a product or service can provide on an online platform. While ratings are quantitative, reviews are expressive. Extracting the users' true sentiments from their review with respect to each aspect is highly insightful. Our project leverages the Stanford CoreNLP Parser to apply PoS tagging, Coreference Resolution and Dependency Inferencing for constructing aspect-sentiment pairs. The aspect polarities are calculated using an amalgam of sentiment lexicons like VADER and TextBlob. We have also used some filtering rules with hard limits to ensure that our system only has the most relevant reviews for processing. One filter obstructs spammed reviews and blacklists the reviewers for the same. Another makes sure that the reviews to be processed have been marked helpful by a strong majority of users. Our system provides recommendations to users based on prioritised aspects. We construct user profiles and product profiles to map according to aspect preferences. The Stanford CoreNLP Parser dependencies have been thoroughly exploited to design the rules for aspect-sentiment extraction.

***Keywords**:- Natural Language Processing, Sentiment analysis, Opinion parsing.*

I. INTRODUCTION

The aspect based recommendation system is still a research topic, and there have been various approaches researchers have undertaken to try and improve the accuracy and thoroughness of the system. Our system is made with a rounded perspective that has the primary motive of giving the most relevant recommendations to the user. The subject is very intriguing, mainly because the instinctive medium that humans use for communication is of words and not numbers. Ratings and reviews are the most common form of feedback that customers of a product or service can provide on an online platform. While ratings are quantitative, reviews are expressive. Extracting the users' true sentiments from their review with respect to each aspect is highly insightful.

These insights are leveraged to provide the businesses with the overall sentiment about a product category in the market. The businesses get a holistic view of the demand of the most preferred aspects with respect to a particular product. Temporal patterns can be observed by businesses with breakthroughs in technology.

For example when the octa-core processor came into picture, the processor aspect of smart phones certainly grabbed the most attention. Similar is the current trend, where the camera aspect is turning heads after the dual camera technology for portraits has been adopted as the signature selling point by a maximum of top smartphone brands.

The facts gathered through our system can lead to some fine grained inferences and implications. Each user, with each purchase or review, makes a contribution in building her profile. It helps the user choose, not the best product in the market with a general viewpoint, but the product that is right for her.

II. LITERATURE REVIEW

The literature survey involved looking for relevant IEEE papers published by students or well renowned authors. These papers played a role in giving us a guideline for our approach toward the problem and an overall depth of the problem and its extent. It also helped us to get an idea about the extent to which the problem has been perused and dealt with. We were also able to define the scope up to which our proposed system will take care of the issue. The challenges faced by the authors of these papers gave insights and will help us to plan mitigation strategies if we face similar challenges in the implementation. The key words or phrases we used to find the most relevant research papers are Natural Language Processing, Sentiment analysis, Opinion parsing, Aspect Extraction, Sentiment Extraction.

In [1], state-of-the-art studies into two principal branches: review-based user profile building and review-based product profile building. In the user profile sub-branch, the reviews are not only used to create term-based profiles, but also to infer or enhance ratings. Opinions with a broad perspective can further be exploited to extract the weight/value preferences that users place on particular features. In [2], a novel approach to introduce aspect-based sentiment analysis into recommender systems is proposed.

The aspect of the product using the topic model is extracted and then the aspect-specific sentiment words are identified using the SentiWordNet (a sentiment lexicon). The use the result of sentiment analysis is then used to make user interests model and the product model. By comparing two models of each user-product pair, we obtain the similarity of the user's interest and the product.

The sentiment analysis system proposed in [6] performs two key functions, aspect extraction and aspect sentiment classification. Aspect extraction has the aim to get the sentiment targets on which some sentiments have been expressed. These targets are usually different aspects of entities (e.g., products or services), which are products in our context. Aspect sentiment classification classifies whether the sentiment expressed on an aspect is positive, neutral, or negative. The main advantage of this new model is the novel additional functionality of providing not only recommendations of items to users, but also recommendations of the most valuable aspects that may enhance user experiences with items.

III. TRADITIONAL AND EXISTING SYSTEM

Traditional Recommendation methods usually focus on utilizing product features obtained from structured behaviour information, which only contains coarse grained user interests. The sentiments in textual reviews are not considered.

IV. PROPOSED SYSTEM

The goals of our new proposed system includes two perspectives viz., customers and businesses. The customer perspective includes recommendation of products based on customer interests, preference based product rating, notification related to new developments based on user interests and user or product profile building. The other perspective includes overall sentiment towards product in market, suggested areas of improvement for product, flexible report generation templates and current trends in the market.

Advancement that we are trying in our system is to provide accuracy by comparing to existing purely feature based recommendation systems, robustness by developing the ability to handle large datasets with minimal slowdown or crash, usability by making user interface minimalistic with well spread-out features having responsive layout and scalability by improving the ability to extend the system to a distributed environment.

Our new proposed system mostly divided into two main components viz., Data Preprocessor and Opinion Parser.

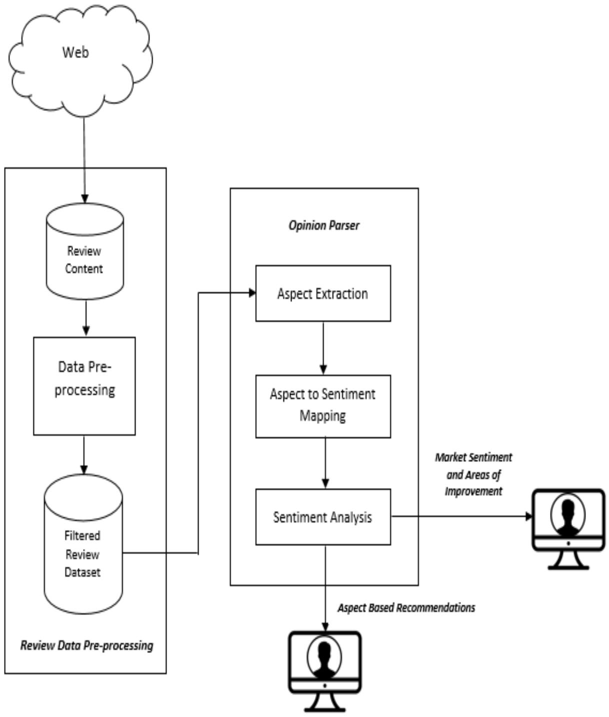


Fig 1:- Block Diagram of Aspect and Review Based Recommendation System

In data preprocessing, There are three steps involved.

A. Check for opinion spamming

- Parse the content of all reviews for the same product.
- Parse the content of reviews for similar selective products.
- If any reviews are found to have exactly matching content, mark all such reviews as irrelevant.
- Set their weightage for the recommendation system to 0.
- Blacklist the spam reviewers.

Example: If the review has the following content: "This phone is really good. Please buy this phone! It's has the best touch screen and the camera is awesome!", this content is compared with reviews on the same product and with the content of reviews on similar products. If the content with other reviews matches, all such reviews are marked irrelevant and their weightage is set to 0, also, the users who wrote these reviews will be added to a blacklist and other reviews by these users will be marked unimportant.

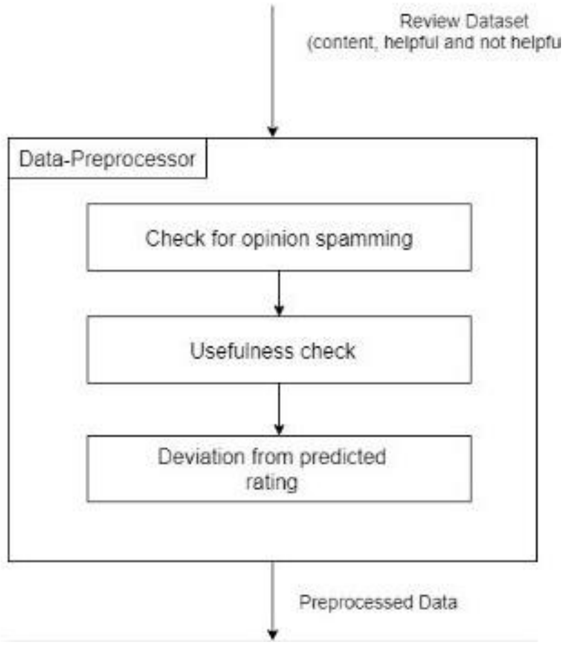


Fig 2:- Data Preprocessor

B. Helpfulness check

- Obtain the number of people who have viewed the review, call it 'v'.
- Obtain the number of people who have found the review useful, call it 'u'.
- Calculate the ratio of u to v.
- Approximate the weight according to this ratio to one decimal place.
- Example: If the review has a high votes-to-views ratio, say a review has been marked helpful by 200 users and not helpful by 150 users, the ratio is 1.33. This number is compared to a threshold number, 0.6, which, if the number fails to exceed, the review is discarded.

C. Deviation from predicted rating

- The adjectives extracted from the review are scored on a scale of 1 to 5.
- An overall expected rating associated with the review is calculated.
- This calculated rating is compared to the actual rating.
- The weight of the review is adjusted on the basis of the deviation of the calculated rating from the actual rating. Example: If a review has all positive sentiments, like "This is the best phone in this price range. The screen is very smooth and responsive. The clarity and vividity of the camera is amazing and the processor can handle heavy apps.", and the entered rating of the phone is 2.5/5, the rating is not in sync with the review content. Such reviews are discarded.

The preprocessed data from data preprocessor is fed as input to the opinion parser. Opinion Parser is a heart of the system. It takes preprocessed data from data preprocessor and produces output which contain recommendations (for users) and overall product sentiment(for businesses). This process takes place in two steps.

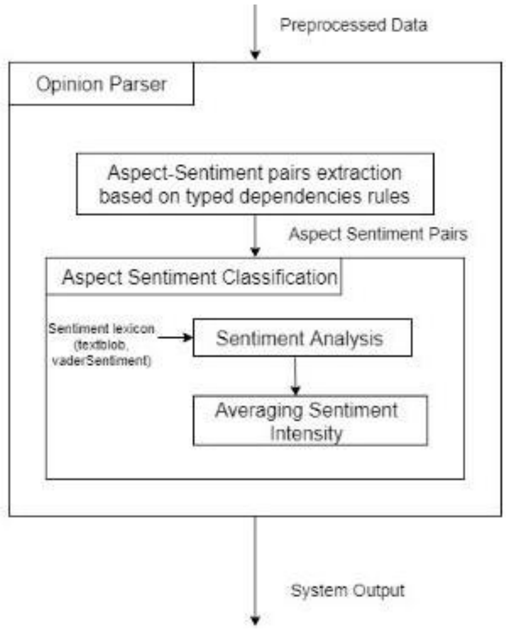


Fig 3:- Opinion Parser

D. Bootstrapping process for aspect extraction

- Bootstrapping is an opinion words and opinion target (aspect) extraction process.\
- In this process a set of opinion words like "good", "bad", "amazing",called as Opinion lexicon, is given as an input to the bootstrapper.
- This Opinion lexicon is used by the bootstrapper to identify opinion words from the reviews.
- Initially bootstrapper uses initially provided Opinion Lexicon to identify opinion words in the reviews. It then extracts corresponding aspects and forms <aspect, sentiment> pairs.
- Known Opinion lexicon and extracted opinion words and target (aspects) are then used together to further extract opinion words and targets.
- Subtasks included in this process are:
  A. extracting targets using opinion words
  B. extracting targets using extracted targets
  C. extracting opinion words using extracted targets
  D. extracting opinion words using both the given and the extracted opinion words.

This process goes on till no opinion words and target are left to be extracted.

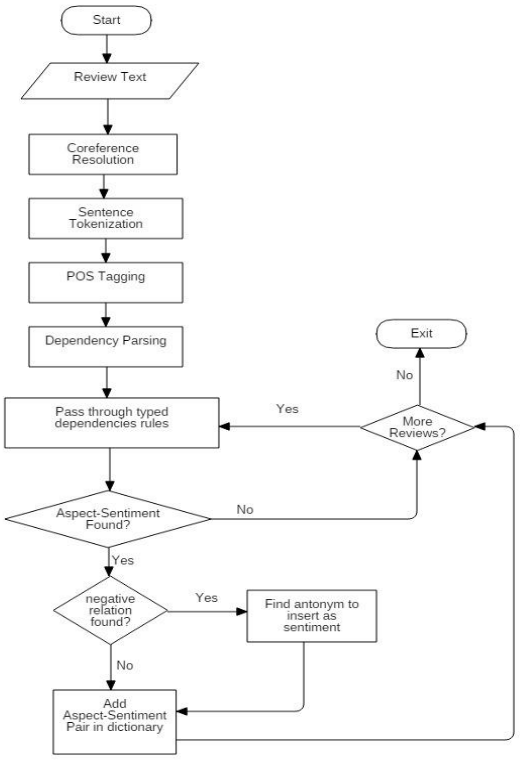


Fig 4:- Bootstrapping Using Double Propagation

E. Aspect Sentiment Classification

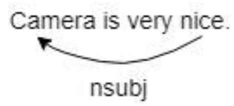
- In this step, intensity of each sentiment of each aspect is calculated using TextBlob and vaderSentimentAnalyazer.
- Intensity is in the range of (-1,+1) with -1 being most negative and +1 being most positive.
- Each aspect is assigned a sentiment score which is average of all sentiment values of sentiments of that aspect.
- Then rating is scaled for each aspect from (-1,+1) to (0,5).

V. ALGORITHM FOR ASPECT-SENTIMENT EXTRACTION

The Stanford typed dependencies representation was designed to provide a simple description of the grammatical relationships in a sentence that can easily be understood and effectively used by people without linguistic expertise who want to extract textual relations. In particular, rather than the phrase structure representations that have long dominated in the computational linguistic community, it represents all sentence relationships uniformly as typed dependency relations. That is, as triples of a relation between pairs of words, such as "the subject of distributes is Bell." Our experience is that this simple, uniform representation is quite accessible to non-linguists thinking about tasks involving information extraction from text and is effective in relation extraction applications.

Typed dependencies rules used for aspect sentiment extraction are.

A. Nominal Subject (nsubj) Rule:

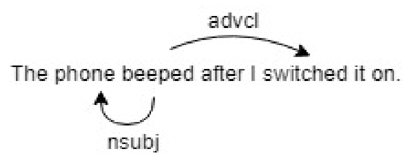


When the sentence has a copular verb, the nsubj dependency has an adjective governor and noun dependent. Hence, here the aspect is the dependent and the sentiment is the governor.

B. Adjectival Modifier (amod) Rule:

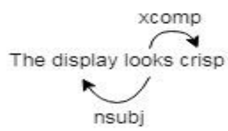
When the sentence has a copular verb, the amod dependency has an noun governor and adjective dependent. Hence here the aspect is the governor and the sentiment is dependent.

C. Adverbial Complement (advcl) Rule:



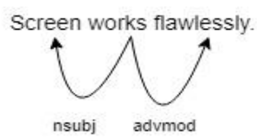
When the governor is a non-copular verb for nsubj dependency as well as advcl dependency, the dependent of nsubj is the aspect and the dependent of advcl is the sentiment.

D. Nominal Subject + complement (acomp/xcomp) Rule:



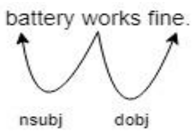
When the governor is a non-copular verb for nsubj dependency as well as xcomp/acomp dependency, the dependent of nsubj is the aspect and the dependent of xcomp/acomp is the sentiment.

E. Adverbial Modifier (advmod) Rule:



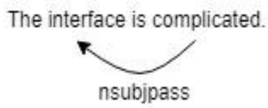
When the governor is a non-copular verb for nsubj dependency as well as advmod dependency, the dependent of nsubj is the aspect and the dependent of advmod is the sentiment.

F. Direct Object (dobj) Rule:



When the governor is a non-copular verb for nsubj dependency as well as dobj dependency, the dependent of nsubj is the aspect and the dependent of dobj is the sentiment.

G. Passive Nominal Subject (nsubjpass) Rule:

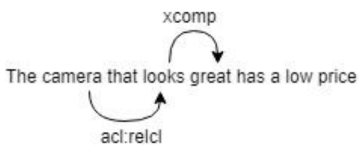


When the sentence has a copular verb, the nsubj dependency has an adjective governor and noun dependent. Hence, here the aspect is the dependent and the sentiment is the governor.

H. Passive Nominal Subject (nsubjpass) + complement (acomp/xcomp) Rule:

When the governor is a non-copular verb for nsubj:pass dependency as well as xcomp/acomp dependency, the dependent of nsubj:pass is the aspect and the dependent of xcomp/acomp is the sentiment.

I. Relative Clause Modifier (acl:relcl) + complement (acomp/xcomp) Rule:

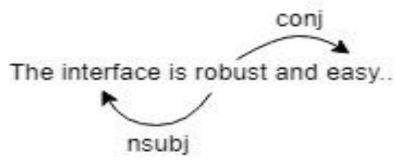


When the dependent is a non-copular verb for acl:relcl dependency as well as xcomp/acomp dependency, the governor of acl:relcl is the aspect and the dependent of xcomp/acomp is the sentiment.

Some rules are such that they need to be associated with each independent rule as they are the most general rules. We have implemented three such rules, namely the conjunction rule, the compound rule for bigrams and the negation rule. These rules are common across the primary rules and their module needs to be implemented as a function only once. This function is called regardless of which rule is triggered by the conditions satisfied. The description of the rules is as follows.

J. Conjunction Rule:

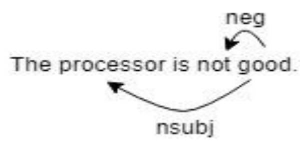
The conjunction rule is applicable when a review text contains multiple sentiment words associated with a single aspect.



In the above example, the conjunction rule is associated with the simple nsubj rule for copular verbs. The basic nsubj rule will only associate the adjective robust with the aspect interface. It will not understand that the adjective easy is also associated with the same aspect. This is where the conj dependency plays a pivotal role. Every word following robust which is separated by commas or conjunctions has a conj dependency with the word robust. In this case, the word easy is associated with robust with the conj dependency. Hence, easy is also associated with the aspect interface as its sentiment.

K. Negation Rule:

The negation rule is necessary because sometimes inverting terms are associated with the sentiment words. In such a case, a word of the opposite meaning of the negated sentiment word is added to the sentiment for the aspect.

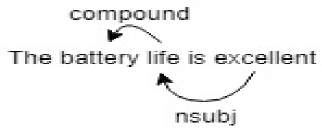


In the above example, the negation rule is associated with the simple nsubj rule. Here, the aspect processor is associated with the sentiment good. Actually, the reviewer has written an inverting term not, before the sentiment word. If it was not for the neg dependency, the aspect-sentiment pair would be processor-good which is completely the opposite of what the reviewer means. The negation rule will help extract the true meaning of what the user has written. It looks for a word closest to the opposite polarity of the extracted sentiment word. This is done by using the getAntonyms() function of the textblob package. In this example, the word having the closest of the opposite polarity of good is bad. Hence, the word bad will be added as the sentiment of the aspect processor.

L. Compound Rule:

If multiple words are playing a role in naming the aspect or sentiment, both the words play a role in constructing the aspect-sentiment lexicon.

ISSN No:-2456-2165



Here, the Compound rule is associated with the simple nsubj rule. The simple nsubj rule extracts the noun life as an aspect and the adjective excellent as the sentiment. The aspect life is completely irrelevant to the product since the reviewer was actually talking about the battery life. Now, the Compound rule checks whether the extracted aspect is the governor for a compound dependency. If such a dependency is found, the dependent of this dependency is appended with the governor, and the bigram is now considered to be an aspect as a whole. Hence, the sentiment is excellent and the aspect is battery life.

VI. COMPARATIVE ANALYSIS WITH EXISTING SYSTEM

Indian e-commerce website, flipkart.com provides aspect specific ratings on a certain limited set of products. We compared how our system fared with the aspect-sentiment extraction system implemented by flipkart. Bold letters in the flipkart reviews indicates sentiment extracted for the aspects.

A. FLIPKART


5★ Really Nice

Pros: 1 : good camera quality in this price segment and front camera is amazing,
2: good bettery life ,
3: good **crispy display**,
4: front side design is pretty good but back side is traditional mi style,
5: powerful processor no lags .
6: face unlock available in OTA update.
Cons: 1: hybrid SIM slot.
2: Ram management was not so good.
Overall this phone is amazing in this price segment.
UDAYAN MONDAL Certified Buyer 24 Feb, 2018

ASPECT BASED RECOMMENDATION SYSTEM
Camera:[“good”]
Display:[“good”,”crispy”]
Processor:[“powerful”]

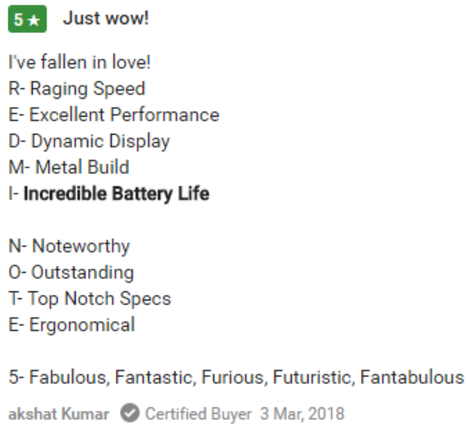
B. FLIPKART


5★ Great product

awesome phone.best camera,ultymate **design.super display.totaly** fully packed devise.
Dinesh P Certified Buyer 23 Mar, 2018

ASPECT BASED RECOMMENDATION SYSTEM
Overall:[“awesome”]
Camera:[“best”]
Display:[“super”]

C. FLIPKART


5★ Just wow!

I've fallen in love!
R- Raging Speed
E- Excellent Performance
D- Dynamic Display
M- Metal Build
I- **Incredible Battery Life**

N- Noteworthy
O- Outstanding
T- Top Notch Specs
E- Ergonomical

5- Fabulous, Fantastic, Furious, Futuristic, Fantabulous
akshat Kumar Certified Buyer 3 Mar, 2018

ASPECT BASED RECOMMENDATION SYSTEM
Processor:[“raging”,”excellent”]
Display:[“dynamic”]
Battery:[“incredible”]
Overall:[“top notch”]

We parsed 50 reviews to measure the accuracy. The accuracy is calculated by aspects extracted/aspects available in review. There were a total of 108 aspects which were relevant to our defined aspects. But our system extracted 92 aspects in all. Therefore, the accuracy is given by (92/108)*100 = 82.14%

The effectiveness of the system is evident by the fact that it uses an all-rounded parser like the StanfordCoreNLP and a well-defined rule set which abides by the typed dependencies of the StanfordCoreNLP parser. We used a dictionary based approach to fetch only the relevant aspects. When the dictionary based approach extracts aspects only based on the dictionary, it covers almost all the relevant aspects, but there is a certain amount of limitation to which the aspects are extracted. Our system works like an expert system and not an intelligent system and hence needs a Knowledge Engineer to make sure that new words or phrases are regularly appended to the dictionary. This is necessary because we have not integrated neural networks or learning algorithms with the system.

The efficiency of our system lies in the fact that the primary focus of our system is on text data. The data structure that we have used primarily is dictionary. Time complexity of finding an element in a dictionary is O(1). We store aspect as the key of the dictionary so all the frequent searches of aspects in aspect-sentiment dictionary are of O(1). Plus, in the algorithm of finding aspect-sentiment pairs we have made a dictionary of dependencies which has dependency name as a key and list of dictionary of governor and dependents as value. It reduces the complexity of finding dependency (to apply appropriate rule) from O(n) to O(1) where ‘n’ is number of dependency relations found for the given sentence(review).

ISSN No:-2456-2165

VII. USER INTERFACE

A. Index Page

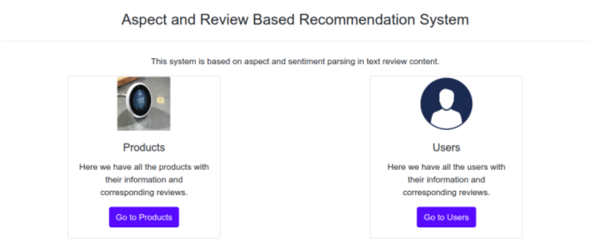


Fig 5:- Index Page

This is the landing page , from where navigation to the list of users and the list of products is possible.

B. Products Index Page

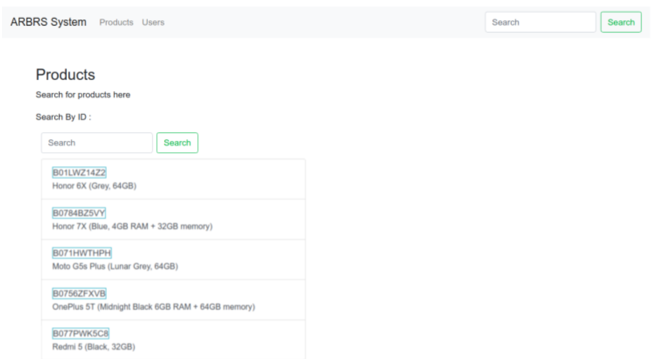


Fig 6:- Products Index Page

This page showcases a list of all the products along with their title and id. Search option is available by using product id as the search field .Clicking on one product takes us to the page specific to that product.

C. Products Page

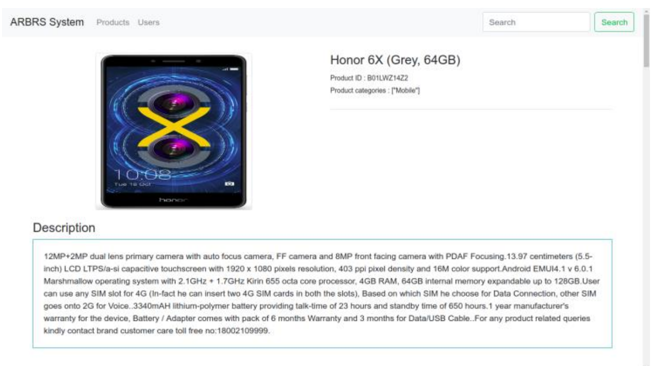


Fig 7:- Products Page

Here an individual product is showcased along with its image , tile , category it belongs to and also its description

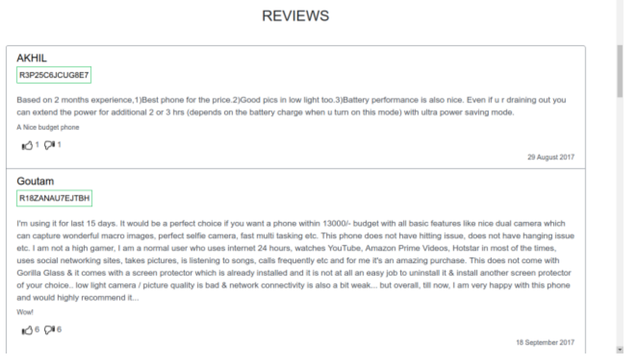


Fig 8:- Products Reviews

All the reviews given for a product are displayed one after the other along with the details of the reviewer with the display of how many people found it useful and how many did not with the help of 'thumbs-up' and 'thumbs-down' glyphicons respectively.

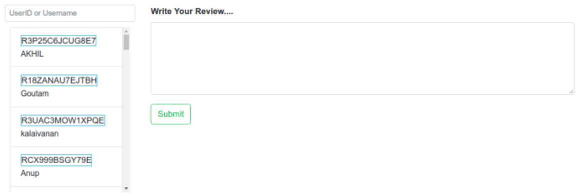


Fig 9:- Write review for products

Under every product is the provision given to provide a review for it.

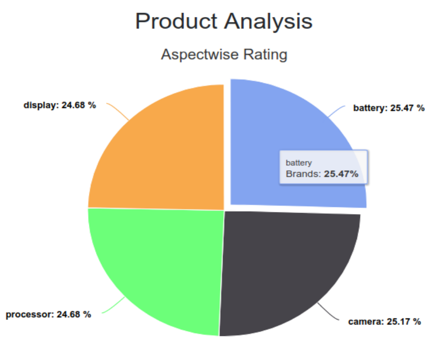


Fig 10:- Product analysis with aspect wise rating pie chart

Here for every product , the each aspect is scaled from a scale of 0-5 to a percentage scale which in essence is the average sentiment of each aspect for each product.

B. User Index Page

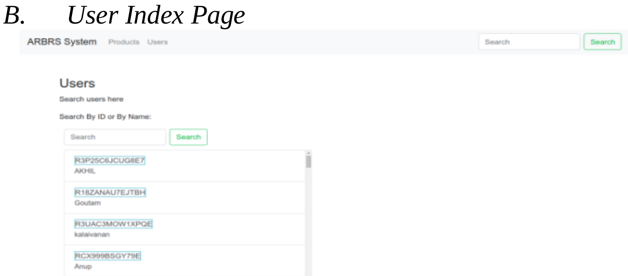


Fig 11:- User Index Page

This is a list of all the users with their names and id along with a search option given which uses user-id as the search field. Clicking on one user takes us to the page specific to that user.

B. User Page

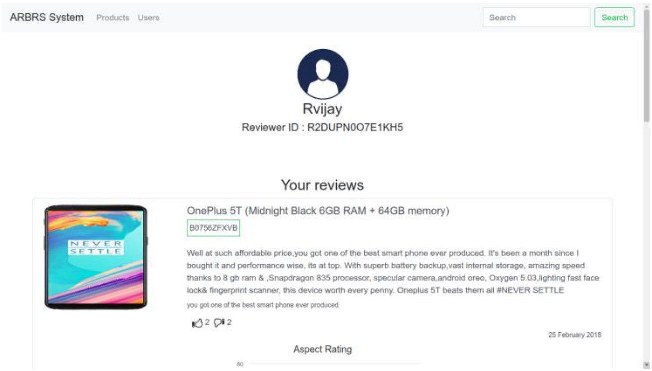


Fig 12:- User page

This page is specific to a user giving details like the user name , user id and a list of all the reviews he/she has given beside the product image for which that particular review is given.

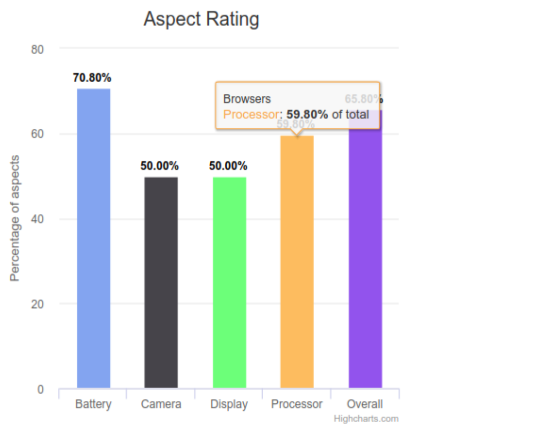


Fig 13:- Aspect Rating of Product

This graph is specific to each review given by the user which gives the user an idea of the overall sentiment of each aspect he/she has talked about in that particular review.

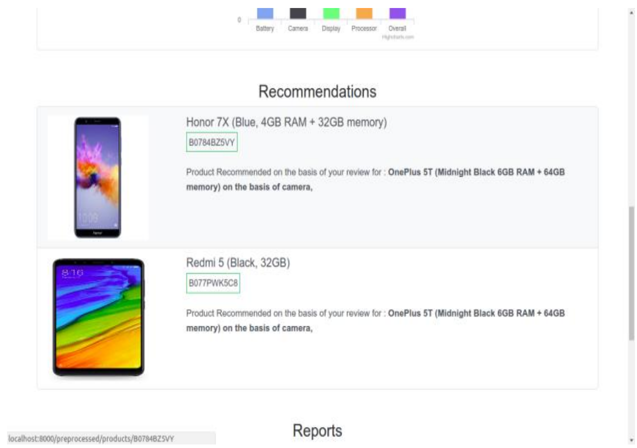


Fig 14:- Recommendation page for users

Each user is given recommendations for various products where clicking on the image of the recommended product takes the user to the page specific to that product.

The user is also given information such as the product recommended to him/her is recommended exactly on the basis of his/her review for which product and which aspect of that product that he/she has talked about.

Fig 15:- Aspect wise Product Recommendation

The graphs give information about the average sentiment rating for each aspect for each product that is recommended to the user by the system.

VIII. CONCLUSION

The project aims at implementing a novel approach in recommendation systems to make them more precise and specific. The system essentially considers a broader set of perspectives which can form the basis for judging whether a person will like a certain product or not and if he/she will, to what extent. If the users of this system who follow the recommendations will rate the product the same or better, it is evident that these users conformed to the recommendations the system provided. The businesses will have a clear picture of the opinion not only about their products, but about specific aspects of their products in the state-of-the-art e-market. This will assist the business in improving the very particular areas where the customers

ISSN No:-2456-2165

are unhappy. It makes the overall business process more efficient and objective. This system is a state-of-the-art system wherein the recommendations are personalized which add to their value.

The rule set is constructed by exploiting the typed dependencies of the Stanford CoreNLP Dependency Parser. For doing this and gain the extent of accuracy that we did, we required to study the Stanford Dependency Manual in depth. The meaning of each typed dependency is explained in this document along with the variety of conditions which trigger for that dependency to show up on the output of the parser. As we constructed direct rules one by one, which were applicable only when certain criteria were met with respect to POS tags and dependencies, we discovered some anomalies, which were relevant to all direct rules. Thus we made the universal rules and associated them with each direct dependency rule. Some other improvements were made like replacing all pronouns with the relevant nouns using the coreference resolver.

This system provides a variety of data in graphical format, which gives different insights to both, buyers and vendors. The buyers are given an output which shows them an aspect-wise comparison for the products best-suited for them. The vendors get an overview of the sentiment of various aspects relevant to a certain product so as to make business plans based on the approach or inclination of the users to a particular aspect. The recommendation system uses the buyer's reviews on various products and which feature of the product the user most frequently mentions in all his/her reviews. This gives the vendors a perspective to focus on targeted advertising and recommendations. The overall sentiment of aspects over a variety of products is also useful so as to give a view of which aspect is considered the most while an average user buying a product and giving the industry a perspective about the universal sentiment.

REFERENCES

[1]. Li Chen, Guanliang Chen and Feng Wang, "Recommender systems based on user reviews: the state of the art".

[2]. Yu Zhang, RuiFang Liu, AoDong, "A Novel Approach to Recommender System Based on Aspect-level Sentiment Analysis".

[3]. Deepali Virmani, Vikrant Malhotra and Ridhi Tyagi, "Aspect Based Sentiment Analysis to Extract Meticulous Opinion Value".

[4]. Niharika Purbey, Aishwarya Iyer, "Categorization and Analysis of Yelp restaurant reviews".

[5]. Warih Maharani, Dwi H. Widyantoro, Masayu Leylia Khodra, "Syntactic pattern based aspect extraction".

[6]. Konstantin Bauman, Bing Liu et al, "Aspect Based Recommendations: Recommending Items with the Most Valuable Aspects Based on User Reviews".

[7]. S.K. Bharti, B. Vachha, R.K. Pradhan, K.S. Babu, S.K. Jena, "Sarcastic sentiment detection in tweets streamed in real time: a big data approach".

[8]. Vaishakh Suresh, Sayeda Roohi et al, "Aspect-Based Opinion Mining and Recommendation System for Restaurant Reviews".

[9]. Sapna Negi, Paul Buitelaar, "Towards the Extraction of Customer-to-Customer Suggestions from Reviews".

[10]. Toqir Ahmad Rana, Yu-N Cheah , "Hybrid Rule-Based Approach for Aspect Extraction and Categorization from Customer Reviews".