A Survey on Product Recommendation and Rating System Based on Product Reviews

PROF.KALAM NARREN
Department of Computer Engineering
Rajarshi Shahu College of Engineering, Tathawade
E-mail: indusharma_25@yahoo.in

.PROF.V.VINAY KRISHNA
Department of Computer Engineering
Rajarshi Shahu College of Engineering, Tathawade
E-mail: kakade.shweta@gmail.com

Dr DINESH KUMAR ashvinimote04@gmail.com,
samanepratiksha@gmail.com, anujayamle@gmail.com Department. of Computer Science and Engineering
Rajarsh Shahu College of Engineering, Tathawade

Abstract— E-commerce website is often used for shopping purposes. In recent times, people have started highly relying on online consumer reviews to know and purchase different products or service. Although few of the recent studies have provided a prediction of the performance of online reviews, this study focuses on the recommendation of user specific products based on reviews which are collected from different e-commerce sites. This generates a need for a system that is capable of classifying the reviews and recommending the most appropriate one. This system consists of two techniques i.e. naïve Bayes for classification and feature ranking algorithm for identifying the features from the reviews. It will not only help the users in buying good quality products but also add enhancement to the e-commerce trade. This study not only solves the problem of humanly analyzing the product before buying it but also add to user satisfaction.

Keywords—feature extraction, sentiment analysis, e-commerce, data analyses, product recommendation

I. Introduction

In our lives, recommendations have a huge significance. People tend to consider something viable if the source that they receive the information from is credible and trustworthy mainly because the advisors have already used those products. They know that people would not recommend a bad product and if they are recommending it, they genuinely mean it. If that be friends, family or online reviews, suggestions and recommendations alter decisions of the buyer. It can be easily seen that shopping websites like Amazon, Myntra, and Snapdeal encourage this sort of review system [1]. This not only makes the customer feel heard but also helps the website in finding out which of their products are preferred and which of them are performing badly. These online forums not only allow individuals to rate their experience regarding their purchased product but also allow them to read reviews left by other people. This new information tends to be a new source of information that helps an individual make a decision of purchase. However, it is very difficult to read all the reviews and then decide whether to buy the product or not. So it is necessary to recommend user specific products by analysing the product reviews. Recommendation systems use collaborative filtering, content based filtering or a combination of both to generate recommendation systems ignores the potential information present in the online reviews. In this paper we poise a system to generate recommendations based on reviews voiced by the customers in different e-commerce platforms. This system has the two following segments, namely feature identification and classification and sentiment analysis [9]. In feature identification, the primary stem is to extract the features of a certain item. Item recommendation is based on sentiment scores of product features. This data mining process has 4 steps carried out one after the other.

These are:

1. Product Availability and extraction:
2. Data pre-processing [2][4]
3. Feature identification
4. Sentiment analysis by computing scores.

Following are the examples of recommender systems: amazon.com, reel.com

Types of recommendation:

1. Personalized recommendation:

Personalized recommendation permits online introduction insertion, data is suggested in any format which is relevant to the user.

2. Non-personalized recommendation:

Non-personalized recommendation system recommends items to the buyers depending on what outer customers have mentioned about the product on an average.

The recommendations do not depend on an individual customer so all customers get similar recommendations.

II. Background:

The way recommendation systems work is that they generally generate a list of recommendations in a few ways i.e. through collaborative content based filtering [6]. Recommendation systems also work with content based algorithm that interact with profiles of users that are generated in the initial stage. In this system we consider different e-commerce sites to collect reviews. As these reviews are from different platforms it provides us with strong training dataset to perform analysis. The recommendation process works in a way of recommending the product as per user interest. The items that fit these criteria are further suggested to the user.

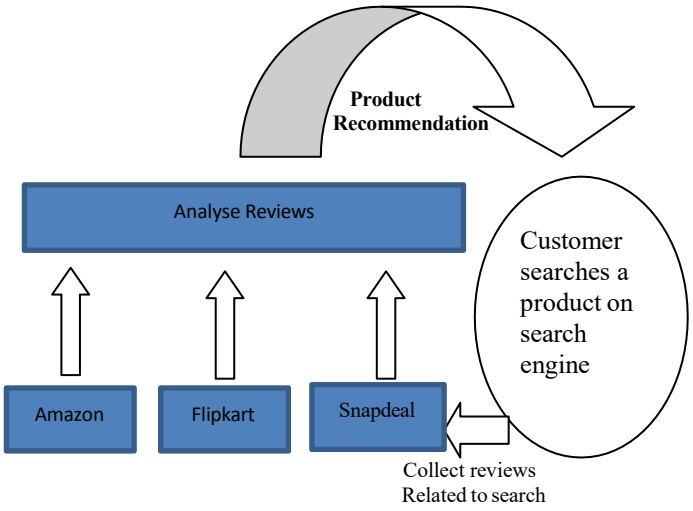

Fig. Flow of Proposed system

In order to classify the reviews to provide recommendations we are using different classifiers.

III. Classifiers/methodology

Classification is a method that assigns the collection to a class. Binary sentiment classification and multi-class sentiment classification are the two approaches of sentiment classification which are often heard. In binary sentiment, every review of the corpus is filed into two categories i.e. positive or negative. Which isn't the case in multi-class sentiment classification where every review can be classified into more than one types? That is very positive or positive or neutral or negative or very negative. Usually binary classification is useful when a comparison of products is in discussion.

For example: if we want to find out the fruit which is long and yellow in colour. In such cases classifiers are required.

1. Naive Bayes (NB) Classifier: It is a classifier which gives us the probability. Decisions are made by analyzing the probability. The main advantage of naïve Bayes classifier is that it does not require large amount of datasets. We only compute the variance of the feature.

The Bayes theorem is given as follows:

$$P(c \mid d) = \frac{P(d \mid c) * P(c)}{P(d)}$$

where

1. P(c) is the hypothesis probability for being true. Also, known as prior probability.
2. P (d) is the evidence probability
3. P (d|c) is the evidence probability for being true
4. P (c|d) is the hypothesis probability when the evidence is there.

How naïve Bayes works:

For example: The camera is **good**.

In the above review the camera is noun whereas good is the feature defining its features. By using naïve Bayes classifier we classify good as positive review. Hence a bag of words is formed consisting positive and negative reviews.

2. Support Vector Machine (SVM) as a classifier: SVM is not a probability classifier. SVM represents each review in the form of vector. This method is used to find the hyper plane represented by $\vec{w}$ .The set of vectors of textual data are said to be separated optimally by hyper plane. Only when it based on two conditions:

1. Separated without error
2. The distance between closest points of each class and hyper plane is maximum.

How SVM works:

E.g.: Condition 1 to find out the right hyper plane in the given dataset:

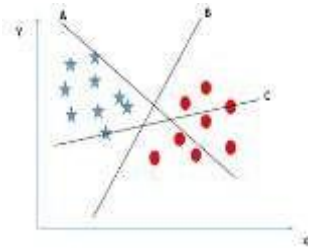

Fig. right hyper plane

3. KNN:

K nearest neighbor's algorithm stores all available cases. It classifies new cases based on a similarity measure (e.g., distance functions).

Statistical estimation and pattern recognition can be done using KNN approach.

KNN is used for classification as well as regression. Widely it is used for classification. The following aspects are considered while evaluating any technique:

1. Interpret output easily.
2. Time required for calculation.
3. Power prediction.

	Logistic Regression	CART	Random Forest	KNN
1. Ease to interpret output	2	3	1	3
2. Calculation time	3	2	1	3
3. Predictive Power	2	2	3	2

How KNN works:

We have to find the nearest neighbour of star(*)

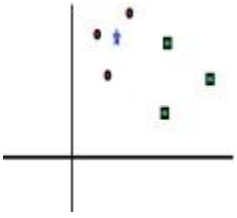

Fig. Finding the class for star

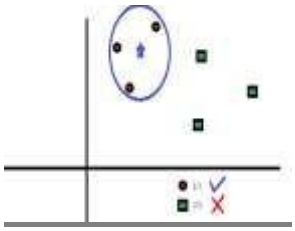

Fig. Nearest data points w.r. to star

We have to find out the class of the blue star. Star can either be Red circle or Green Square or nothing else. Nearest neighbour is denoted as "K" .Let K = 3. Make a circle with star as centre with three nearest data points on the plane. From this we can conclude that all the nearest data points with respect to star are red circles. So star belongs to red circles group. However, it is very important to decide the value of "K".

IV. Recommendation Techniques:

1. Collaborative filtering: Based on user's previous searches. It gives a rating system to the user based on the information collected [6].
Collaborative filtering consists of 3 models: user-based, item-base and model-based
Advantages: Recommends different products that do not satisfy user interest.
Disadvantages: The recommendations are based on the historic data.

2. Content-based filtering: It considers two parameters for recommendation of product [11].
1. Description of the item
2. User interest

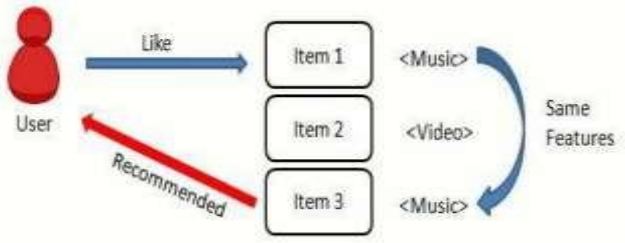

Fig. Content based filtering

Advantages:
1. Data of other users is not required.
2. Recommends new and popular products.
Disadvantages:
1. Requires content analysis.
2. It does not determine the quality of the product
We are using naïve Bayes classification for classifying reviews in proposed system

V. Feature identification algorithms:

Feature identification is a way of uniquely identifying the specification of the given text. The different feature identification algorithms are classified into [12]:
1. **Heuristic based**: A heuristic value is used along with the functional value and then the identification is done.
2. **Statistical:** It is used to perform mathematical operations to identify the features.
3. **Clustering:** We can extract only major features using clustering. However there may be requirement of including the minor features also but this cannot be achieved using clustering
4. **Hybrid:** It includes pos tagging along with wordNet tool. We are using hybrid algorithm for feature identification in proposed system.

VI. Issues in recommendation

The different issues in recommendation systems are [13]:
1. Data collection:
The data which is used by recommendation system can be termed as explicit and implicit data. The explicit data is all the information that the user fills in by him. The implicit data, specifically in e-commerce websites is a transaction data which includes the data of the purchase. Whereas the implicate data hast to be first analysed before it is put to describe user-item ratings.
2. Cold start
Cold start is a problem which occurs when there is no ample data in the beginning stage. The recommendation system falls short of data to produce appropriate recommendations. There are two types of cold start problems. Namely:
1. New user problem and
2. New item problem.
3. Stability v plasticity:
The opposite if cold start problem is stability v plasticity problem. When the user rates lots of products then h=it becomes very difficult to change their preference in the existing user profile.
4. Performance v scalability:
The e-commerce websites should be capable of delivering real time information to the user, millions of data sets have to be scanned in a fraction of time and hence performance is a key point for any recommendation system.
5. Scarcity:
The most active users too can only rate a limited amount of items considering the large volume of the database. As a result of this, even the most popular items have few ratings.
6. Scalability:
Immense computation capacity is essential as at a point of time millions of users are searching for millions of products.
7. Privacy
Privacy is undoubtedly a vital aspect; any recommendation system cannot compromise privacy of any user.

VII. Conclusion and future work:

Recommendation systems are useful for the user to avoid unwanted excessive information. The e-commerce websites grow at a rapid rate and the system must be capable of keeping up with that growth. These systems are beneficial for customers to find their products efficiently while on the other hand they also benefit the providers to enhance their sales. The best example of recommendation systems can be amazon.com and YouTube where one system is for products and the other is for videos respectively. In the upcoming years we can definitely see improvements and modifications in this system so this system can be used for wider applications

VIII. References

[1] A Review Paper on E-Commerce.
[2] The Impact of applying Different Preprocessing Steps on review spam detection
[3] A Credibility Analysis System for Assessing Information on Twitter
[4] A Comparison between Preprocessing Techniques for Sentiment

[5] Analysis in Twitter
[6] User specific product recommendation and rating system by performing sentiment analysis
[7] E-commerce Product Recommendation Method based on Collaborative FilteringTechnology
[8] Sentence-level sentiment polarity calculation for customer reviews by considering complex sentimental structures
[9] Assigning Polarity Scores to Reviews Using Machine Learning TechniquesSentiment analysis using product review data
[10] Sentiment analysis of product reviews for e-commerce recommendation
[11] Feature-based Customer Review Mining
[12] A Review of Feature Extraction in Sentiment Analysis
[13] Recommender Systems: Issues, Challenges, and Research Opportunities