E-Commerce - Bookstore with Recommendation System using Prediction

Ashwathy.R¹, Aswitha.A²
¹, ² UG Student, Department of Computer Science and Engineering
Sathyabama University, Chennai- 600119, Tamil Nadu, India.

Nagarajan.E³
³ Assistant professor, Department of Computer Science and Engineering
Sathyabama University, Chennai- 600119, Tamil Nadu, India.

Abstract

E-commerce is a platform to buy and sell products electronically via the Internet. E-commerce channels like Amazon.com, Olx, Flipkart are gaining wide spread popularity. In a typical bookstore, users are often presented with the recommendation which is not relevant to their purchases. The proposed system aims at solving the issue by implementing a prediction based algorithm [1] and a feedback based algorithm [2] which enhances the already existing recommendation system. The main goals of this bookstore project are to get a firm understanding of the e-commerce architecture and bookstore application, determine the pros of including features such as feedback submission, predicting customer preferences from past purchases and feedback data for making suggestions. The system is designed to support high accuracy, design flexibility and improve availability.

Keywords: Prediction; Recommendation

I. INTRODUCTION

Life of a typical human has been fundamentally redefined by the use of information technology and services. Internet has changed the way we communicate with each
other, how information is distributed across the globe etc. E-commerce has revolutionized the way the world does trade and commerce using information and technology services. The increase in the flexibility of doing business anytime from anywhere in the world has also given rise to increased use of e-commerce compared to a physical buys and sell model. People are much more comfortable doing trade through the use of e-commerce rather than having a physical presence at the point of sale. Any e-commerce system facilitates faster buying / selling of products, 24*7 availability, low operational cost and improved quality of services. The integration of social media with the other e-commerce sites has been done in abundance these days [7]. Even though online recommendation of products has been examined before [3], the study focuses only on developing results within e-commerce sites and mainly uses users’ transaction record history. The challenge that is faced by such websites are that no prediction is made on the users’ particular interest on products they buy. A random set of similar items are given as recommended products. No effort is made to derive the users’ interest. To address this particular challenge, an approach is proposed with the use of certain algorithms, one for prediction and one for recommendation. This approach proposes the use of prediction algorithm like Gradient Boosting [1] which will analyze and predict the users’ interest. It also proposes an altered gradient boosting tree designed to modify users’ product attributes to potential feature representation that can be used for recommendation of products. This approach is modelled to use an approach based on clustering which is the self constructing [2] algorithm to reduce the magnitude of the gathered data.

II. RELATED WORKS
In the system proposed [1] an incorporation of social media sites with an e-commerce website is done. It included the micro blogging information from a huge dataset which was used to analyze the activity of the customers. In the approach proposed [2], a new system of recommendation is introduced. The main aim of the given system is to establish a clustering technique in order to group the related products and to minimize the dimensionality. Similar approach has been given in [9]. The main aim of this approach is also to cluster the datasets into similar categories and apply a set of self-constructing algorithm to reduce the size of the dataset. In the proposed approach [7], a combination of both prediction and analysis of social media has been depicted. This approach helps to analyze the various methods used to perform prediction on various sites. In the proposed system [5], the behavior of the purchases done is examined.

III. METHODOLOGY
The following processes are the modules and the methodologies that are followed in our proposed system to get a effective recommendation and prediction process for the E-Commerce bookstore.
3.1 Login and Registration
The login and registration page is set up for the users’ to access the bookstore application. The users’ are provided with a page where the login is made mandatory. If users’ have not logged in already, they are made to register first. For registration, details like username, password and mail-id are to be given. On successful registration, the page forwards itself to the login page. All the data are stored in a database. Whenever users’ who have not registered first wishes to check out, they are provided with the login page where they are made to register.

3.2 E-Commerce System
The E-commerce system fig(1) is a user interface designed to make an easy access to the application by users’. The UI is a collection of books that are made into various categories. The collections of books are taken from various datasets like Amazon. The e-commerce system generally has a home page which displays all the books and the various categories displayed. It also includes a search box and a list of upcoming books. The entire purchase of books will be shown in terms of numbers in the shopping cart. Every book that users’ select will be provided with a set of descriptions. Users’ can add a particular book to the cart with the help of add to cart option. As soon as users’ add the book, the page will get redirected to the page with the purchase details where all the details will be displayed including the total amount to be paid. Once the books are added, users’ can checkout with the different payment methods provided.
options. A contact button is also included on the home page where users’ can give feedback regarding the use of the site. Most approaches do not include information from other sites into the e-commerce site [4],[5]. Our approach solves this issue.

3.3 Crud
The CRUD (Create, Retrieval, Update, and Delete) generally performs all of the modification activities. Create - allows users’ to create an account to access the application. Every account that is created will be stored in the database. On successful registration, all of the details given during the registration will be stored in separate fields in the database. Retrieval – used to retrieve any account that has been added to the database. When users’ visits the site for the second time, on successful login, that particular user’s account is retrieved from the database. Once logged in all previous purchases made by users’ will be displayed. Update- Any changes made by users’ on the account will be updated inside the database using this operation. The admin will have a track of all the updates taking place in the database. Delete- If users’ wishes to delete the account it can be done so. On deletion, the complete account of the user will be deleted from the database. The deleted users’ account will not be visible in the application again. Any deleted user will need to register to use the application again.

3.4 Book Search Engine
The bookstore application is provided with a search box to search for any books. As users’ search for a particular book, a search will be done on categories, description and the title of the book. The search keyword will be queried in the database and if a match is found, the results are displayed on the screen. If the result is not found in the database then an alert will be displayed. Search is usually done to make the work of users’ easy. Users’ need not browse through the categories and search for a particular book instead type a single word of the book and get the search result of that particular book.

3.5 Analysis and Recommendation System
The analysis engine is designed specifically to analyze the users’ action in the application. If users’ log into the account more than two times, during the third time users’ are presented with book set of their interest. An analysis is made on the users’ history of purchase. Every book that the user purchases will be stored in the database with the help of book details. The recommendation system is thereby used to recommend the books of users’ interest to the user. Certain algorithms are used to both predict as well as to recommend books to users’. The content provided by the user is administered by the content recommendation system [8].

3.5.1 Algorithm Implementation
(I) Self Constructing Clustering Algorithm:
The purpose of self constructing cluster algorithm is to group the pattern into a collection of clusters, identical patterns are grouped into similar clusters and non-identical patterns are grouped into different clusters. The user need not know the
working of clusters. It is not necessary to create the cluster at the start. The cluster can be created during the process implementation [2] [9].

Let a set $A$ of $x$ patterns $a_1, a_2, \ldots, a_x$, with $a_i = A_{i1}, A_{i2}, \ldots, A_{ip}$ for $1 \leq i \leq x$. Let $R$ be the number of currently existing clusters. The clusters are $E_1, E_2, E_3, \ldots, E_R$, respectively. Each cluster $E_j$ has mean $m_j = m_{j1}, m_{j2}, m_{j3}, \ldots, m_{jp}$ and deviation $\sigma_j \leq \sigma_{j1}, \sigma_{j2}, \ldots, \sigma_{jp}$. Let $S_j$ be the size of cluster $E_j$. Initially, we have $R = 0$ [6]. So, no clusters exist at the beginning. For each pattern $a_i$, $1 \leq i \leq x$, We have to calculate membership degree of $a_i$ in existing cluster,

$$
\mu_{E_j}(a_i) = \prod_{q=1}^{p} \left[ 1 - \left( \frac{A_{iq} - m_{jq}}{\sigma_{jq}} \right)^2 \right] \quad (1)
$$

For $1 \leq j \leq R$

In eqn (1), $a_i$ passes the similarity test on cluster $E_j$ if $\mu_{E_j}(a_i) \geq \rho$.

Where $\rho$ is a predefined threshold. Two cases may occur. In the first case, there are no existing clusters in the pattern. In this case we create new cluster $E_r$, $r = R + 1$ is created with,

$$
m_r = a_i, \sigma_r = \sigma_0 \quad (2)
$$

Whenever the new cluster is created, the number of clusters is increased by 1 every time. The size of cluster $S_r$ is initialized to 1.

$S_r = 1, R = r$. In the second case, if a pattern is matched to already existing cluster. Let $E_v$ be the highest membership degree,

i.e

$$
v = \max_{1 \leq j \leq R} \mu_{E_j}(a_i) \quad (3)
$$

In this case (3), $a_i$ is most similar to cluster $E_v$, and $m_v$ and $\sigma_v$ of cluster $E_v$ should be modified to include $a_i$ as its member. The modification to cluster $E_v$ is described as follows:

$$
\sigma_v = \sqrt{M - N \cdot \sigma_0} \quad (4)
$$

$$
M = \frac{(S_v - 1)(\sigma_{jq} - \sigma_0)^2 + S_v \sum_{j=1}^{p} (A_{jq} - m_{jq})^2}{S_v} \quad (5)
$$

$$
N = \frac{S_v + 1}{S_v} \left( \frac{S_v \sum_{j=1}^{p} (A_{jq} - m_{jq})^2}{S_v + 1} \right) \quad (6)
$$

$$
m_v = \frac{S_v \sum_{j=1}^{p} (A_{jq} - m_{jq})}{S_v + 1} \quad (7)
$$

for $1 \leq j \leq p$, and

$$
S_v = S_v + 1 \quad (8)
$$
The value of R does not change. Process iteration is done until all patterns have been processed. For every A, we will get R clusters.

\[
\begin{align*}
\text{of original pattern of words:} & \quad n \\
\text{of classes:} & \quad q \\
\text{Initial deviation:} & \quad \sigma_1 \\
\text{Initial } S \text{ of clusters:} & \quad i = 0 \\
\text{Input: } y_j = \langle y_{j1}, y_{j2}, \ldots, y_{jq} \rangle, 1 \leq j \leq n \\
\text{Output: Clusters } H_1, H_2, \ldots, H_i \\
\text{Procedure : Self-Constructing-Clustering-Algorithm} \\
\text{for each word pattern } y_j, 1 \leq j \leq n \\
\text{temp } Z = \{ H_t | \mu_{H_t}(y_j) \geq q, 1 \leq t \leq i \} \\
\text{if temp } Z = \emptyset \\
\text{A new cluster } H_u, u = i + 1, \text{ is created} \\
\text{else let } H_b \in \text{temp } Z \quad // \text{be the cluster to which } y_j \text{ is closest} \\
\text{Incorporate } y_j \text{ into } H_b \\
\text{end if} \\
\text{end for} \\
\text{return} \quad // \text{with the created } i \text{ clusters}; \\
\text{end procedure}
\end{align*}
\]

This algorithm depicts the use of clustering in order to divide every data and analyze them one by one.

\[\text{(II) Gradient Boosting Algorithm:}\]
The purpose of the gradient boosting algorithm is to predict the activity of the user. The term Boosting refers to the method of prediction. Generally, this algorithm has several parts. We are using the gradient trees. They consist of weak learners. When you want to perform a task, you check for the different ways in which it can be done. If it cannot be done in the given ways, then it is called as the weak learner in gradient boosting. The gradient trees are mainly used for decision making. Tree Boosting is usually a set of decision trees. It together forms a prediction model. Every time data is added to the tree, the tree grows and computes its accuracy.

\[
\begin{align*}
\text{Procedure: Gradient Boosting Algorithm} \\
\text{Input: } \{ S = S1 = S2 = 0 \quad // \text{Weak learner} \\
\text{alpha} = 0, x, y, u \} \\
\text{Output: } P = P + \alpha \quad // \text{Predict value} \\
\text{Initialize } i \\
\text{for each } i = 1 \text{ to } n \\
R = (y, x) (y, x) \quad // \text{Re-weight}
\end{align*}
\]
The above algorithm gives a better understanding of gradient boosting technique.

IV. CONCLUSION
In the proposed system, the given algorithm is assumed to examine the problem of prediction and recommendation as a part of an e-commerce website. The goal is to provide users’ with a set of recommendations on their particular product of interest with a set of predictions made. Any type of predictions can be made using the specified algorithms. There are a lot of products in every e-commerce website and performing recommendation is a huge task. The given approach reduces the dimensionality of the products thereby making the recommendation easier. Further improvements can be made on the given algorithms for better efficiency.

REFERENCES

