Machine Learning Techniques with Recommender System

Mrs. Zinkal Patel, Assistant Professor, ROFEL BBA & BCA College, Vapi, India,

zinkal.patel121@gmailcom

Mr. Navin Deepakbhai Bhanushalee, Assistant Professor, ROFEL BBA & BCA College, Vapi,

India, navindbhanushali9033@gmail.com

ABSTRACT - Recommender systems are machine learning techniques to represent relationship between product and those are design to recommend product to the user based on many different factors like historical data, reviews etc. In this review paper I describe several machine learning techniques including Linear Regression, Logistic Regression, Decision Tree, SVM, Naïve Bayes, KNN. Those are used in previously commonly used recommender systems. These techniques can be applied on different pattern retrieved using different data mining techniques and through which we can predict the future relationship between product and user. He we propose the platform to address the problem of suggesting most suitable item components for each user at every time by creating a recommender system components using machine learning techniques. Once we have gathered all the data of the recommender system then we can find a pattern and on that pattern applying different machine learning techniques to determine future prediction about product and user relationship. Here we discussed all the machine learning techniques and methods applied on different food recommender system to retrieve future relationship between user-items.

Keywords: decision tree, linear regression, logistic regression, machine learning techniques, Naïve Bayes, recommender system

I.   INTRODUCTION

During the last few decades the use of YouTube, Netflix, Amazon and many other such web services and other recommender systems have taken more and more place in our lives. So recommender systems are today unavoidable in our daily online journey. In a general way, recommender systems are algorithms aimed at suggesting relevant product to the users. Recommender systems are really critical because they can generate a huge amount of income. Recommender systems are the systems that are designed to recommend things to the used that is based on many different factors line history etc. these system predict the most product that the users are most likely to purchase. It help the users to identify the correct product for them.

II.   LITERATURE REVIEW

Shina, Shikha Sharma and Anshu Singla (2018) conducted study on "A study if Tree Based Machine learning Techniques for restaurant Reviews". The objective of this survey includes classifying restaurant into several classes based on their service parameter and apply popular machine learning algorithms like decision tree and random forest on dataset of 8500 records to find which method gives best result for prediction of the class of restaurant.

The result is that the decision classifier is more effective than random forest[1].

Assistant Professor. J. Priya(2020) conducted a prediction on "Predicting Restaurant Rating using Machine Learning and comparison of Regression Models". The objective of these survey was to build a model using various regression algorithms and with the help of these model, restaurants decide their menu, cuisine, theme, cost, demographic location etc. and also helps to understand the factor affecting the rating of restaurant. Here database for restaurants identified, after that data mining techniques are applied to identify pattern and of the food culture[2].

Dr. Kali Charan Modak and Dr. kumkum Sinha(2019) developed a customer satisfaction model for online food delivery services by using logistic regression. The objective of this research is to develop a logistic regression model of customer satisfaction for Online Food Delivery in Indore region[3].

Mona Mishra and Yifan Gong(2018) study on "Recipe Recommender System Using Image Recognition of Food Ingredients". The objective of this was to build a recipe recommendation system employing image recognition of food ingredients, specifically vegetables. First they build a convolutional neural network (CNN) to recognize food

image and predict what the object in the image is and then use Support Vector Machine to classify them[4].

Mansi Goel, Ayush Agarwal, Deepak Thukral, Tanmoy Chakraborty(2019) present "Fiducua: A Personalized Food Recommender System for Zomato". The objective of this was to identify item-user related information in review and accordingly suggest an appropriate restaurant for the particular item depending on user-item and item-item similarity. Here Naïve Bayes technique use and it give the probability of positive and negative sentiment for the item[5].

Zainur Rimadhon, Eko Sediyono, Catur Edi Widodo(2020) presents "Various implementation of collaborative filtering-based approach on recommendation systems using similarity". The objective to implement this recommendation system provides recommendations to users who want to order or choose fast food menus and provide recommendations based on item data calculations with customer review data using a collaborative filtering approach[6].

A. Prabhat and V. Khullar(2017) presents "sentiments classification on big data using naïve bayes and logistic regression". In this paper, twitter reviews have been used using machine learning algorithm line naïve bayes and logistic regression for ckassification task. In this the analysis with logistics regression gives 10.1% more accurate and 4.34% more precise results with almost one fifth implementation time for same size of dataset[7].

Goel, Mansi,Agarwal, Ayush,Thukral, Deepak Chakraborty, Tanmoy(2019) presents "Fiducia: A Personalized Food Recommender System for Zomato". In this paper Fiducia, a food review system involving a pipeline which processes restaurant-related reviews obtained from Zomato (India's largest restaurant search and discovery service). Fiducia is specific to popular café food items and manages to iden- tify relevant information pertaining to each item separately in the reviews. It uses a sentiment check on these pieces of text and ac- cordingly suggests an appropriate restaurant for the particular item depending on user-item and item-item similarity. Experimental re- sults show that the sentiment analyzer module of Fiducia achieves an accuracy ofover 85% and our final recommender system achieves an RMSE about 1.01 beating other baselines[8].

III.   METHODS

Decision Tree[1]

Decision Tree algorithm is classification technique due to its human-friendly approach and close correspondence to the real time trees. In this algorithm the decisions are taken based on the traversal of tree that depending upon the various parameters. The first node of the tree is known as the "root" node. The tree moves downwards the "internal"

nodes to the "leaf" node. The leaf nodes are the ones which do not have any downward successor. The decision of placement of nodes is taken on the basis of some mathematical calculations of "Information Gain" and that is based on the Entropy that bequeaths the capability of consistency of an attribute. The higher information gain values, the more chances of becoming the root node of attribute. The data is split further and information gain is calculated for each attribute. One of major flaw in this methodology is inconsistency of decision tree which leads change to change in whole structure of tree on occurrence of a small change in the data.
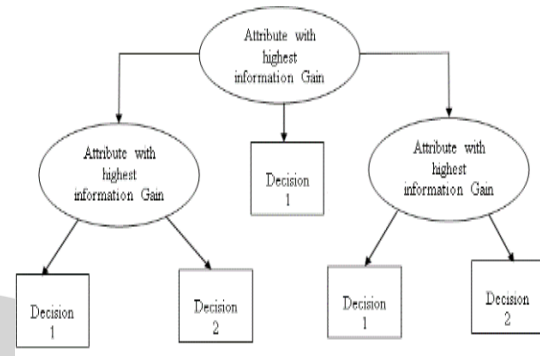

Figure 1: Decision Tree

Each node of the tree depicts an attribute and the branches are available options to choose from as shown in following figure 1 that is of decision tree.

In that we can put attributes with highest information gain at root and after that we can take decision. Last node is known as child node that is decision node that is describe in figure 1 decision tree.

Linear Regression Algorithm[2]

Linear regression algorithm is the simplest regression technique. The purpose is to build a functional model(relationship) between dependent variable(predicted variable) and independent variable(exploratory variable). There are simple linear regression and multilinear. The simple linear regression technique have only dependent and one independent variable while multilinear regression consists of multiple dependent and multiple independent variables.

A linear regression line can be represented as,

$$Y = a + bX,$$

Where X is independent variable.

Y is dependent variable.

B represents the slope of the line and a is the intercept (the value of Y when X=0)

Logistic Regression[5]

Logistic Regression is classification algorithm so it is used to deal a classification problem. It gives the binominal outcome as it gives the probability if an event will occur or

not (in terms of 0 and 1) based on values of input variables. For example, predicting if a tumour is malignant or benign which can be considered as binominal outcome of Logistic Regression. There can be multinomial outcome of Logistic Regression also. E.g. prediction of type of cuisine preferred: Punjabi, Chinese, Italian, Mexican etc. There can be ordinal result also: product rating 1 to 5 etc. So Logistic Regression technique deals with prediction of target variable and that is categorical.

Logistic Regression has the following advantages: simplicity of implementation, computational efficiency, efficiency from training perspective, ease of regulation. No scaling is required for input features. Logistic Regression is a probability score so to apply it for solving business problem it is required to specify customized performance metrics so as to obtain a cut-off which can be used to do the classification of the target. Also Logistic Regression method is not affected by small noise in the data. Main disadvantage of this is inability to solve non-linear problem.

SVM (Support Vector Machine)[5][9]

Support Vector Machine (SVM) can handle both classification and regression problems.in this method hyperplane needs to be defined which the decision boundary is. When there are a set of objects of different classes then decision plane is needed to separate them. The objects may be linearly separable or may not be linearly separable in that case complex mathematical functions called kernels are needed to separate the objects which are members of different classes. SVM method mail objective is to correctly classifying the objects based on examples in the training data set. SVM method can handle both semi structured and structured data, it can handle complex function if the appropriate kernel function can be derived and it can scale up with high dimensional data. But the disadvantages of SVM are, it performance goes down with large data set due to the increase in the training time. SVM does not work well when dataset is noisy. SVM does not provide probability estimates.

Naive Bayes[7][8]

Naive Bayes classifier is a probabilistic classifier which is refers to Bayes Theorem. NBC is based on conditional probabilities with independent assumption between its features. The Naïve Bayes classifier assumes categorical class labels and categorizes data based on training set and values in test data. Applications of Naive Bayes technique are spam filtering, text classification, Hybrid Recommender System. Mathematically Naive Bayes classifier is defined as following equation 1:

$$P(X|E_1,\ldots,E_n) = \frac{P(E_1,\ldots E_n|X)P(X)}{P(E_1,\ldots,E_n)}$$

Equation 1

Where, X is the probability of an event.

E is the given evidence.

$P(E_{1\ldots}E_n \mid X)$ = Likelihood.

$P(X)$ = Prior.

$P(E_1\ldots E_n)$ = Normalization Constant.

K-Nearest Neigbour (KNN)[6][10]

One of the filtering type of recommender system is collaborative filtering. In collaborative filtering one of the algorithm used is the K-NN algorithm that is also known as k nearest neighbors algorithm which is the most commonly used classification algorithm. K-Nearest Neighbor is a method used to classify new objects based on attributes and training samples that have been classified. This algorithm is used to solve various category problems. K-Nearest Neighbor is included in the lazy learning algorithm which is easy to implement.

The steps of the KNN algorithm:

i. Determine the parameter k that is number of nearest neighbor.
ii. Find distance of training data with all test data.
iii. Sort the distance based on the smallest value of k.
iv. Determine the group of test data that is based on k.

Calculation of metric distance on the KNN algorithm we can use cosine similarity, cosine similarity is a method for measuring the level of similarity between two vectors. Calculation in this methods are done by calculating the Cosine value between two vectors. The proximity of user characteristics can be found by cosine similarity formula using the following Equation 2.

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \, x \, B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \, x \, \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Equation 2.

Where,

A = document vector

B = query vector

A . B = multiplication between vector A and vector B

$\|A\|$ = length of vector A

$\|B\|$ = length of vector B

$\|A\| \, \|B\|$ = multiplication items between $\|A\|$ and $\|B\|$

IV.   CONCLUSION

The recommendation system is using a collaborative filtering approach that provide a fairly item recommendation. By using different machine learning algorithm on past review, ratings, past orders and user

information, we can increase similarity in recommendation and improve user based recommendations or item user recommendation. It is important to have more reviews from many users so we can improve the accuracy of recommendation such as type of food, price etc. so in recommender system by applying different machine learning algorithm the results regarding relationship can be obtained.

REFERENCES

[1] Shina, S. Sharma, and A. Singla, "A study of tree based machine learning techniques for restaurant reviews," *2018 4th Int. Conf. Comput. Commun. Autom. ICCCA 2018*, pp. 1–4, 2018, doi: 10.1109/CCAA.2018.8777649.

[2] J. Priya, "Predicting Restaurant Rating using Machine Learning and comparison of Regression Models," *Int. Conf. Emerg. Trends Inf. Technol. Eng. ic-ETITE 2020*, pp. 1–5, 2020, doi: 10.1109/ic-ETITE47903.2020.238.

[3] K. C. Modak, "A Logistic Regression Model of Customer Satisfaction for Online Food Delivery Services," *UNNAYA Int. Bul. Manag. Econ.*, no. 2349=7165, pp. 320–326, 2019.

[4] M. Mishra and Y. Gong, "Recipe Recommender System Using Image Recognition of Food Ingredients," 2018, [Online]. Available: www.aaai.org.

[5] S. Ray, "A Quick Review of Machine Learning Algorithms," *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019*, pp. 35–39, 2019, doi: 10.1109/COMITCon.2019.8862451.

[6] Z. Romadhon, E. Sediyono, and C. E. Widodo, "Various Implementation of Collaborative Filtering-Based Approach on Recommendation Systems using Similarity," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 3, pp. 179–186, 2020, doi: 10.22219/kinetik.v5i3.1062.

[7] A. Prabhat and V. Khullar, "Sentiment classification on big data using Naïve bayes and logistic regression," *2017 Int. Conf. Comput. Commun. Informatics, ICCCI 2017*, no. January 2017, 2017, doi: 10.1109/ICCCI.2017.8117734.

[8] M. Goel, A. Agarwal, D. Thukral, and T. Chakraborty, "Fiducia: A Personalized Food Recommender System for Zomato," no. c, 2019, [Online]. Available: http://arxiv.org/abs/1903.10117.

[9] D. Barbella, S. Benzaid, J. Christensen, B. Jackson, X. V. Qin, and D. Musicant, "Understanding Support Vector Machine Classifications via a Recommender System-Like Approach," *Proc. Int. Conf. Data Min.*, pp. 305–11, 2009.

[10] M. Premasundari and C. Yamini, "Food and Therapy Recommendation System for Autistic Syndrome using Machine Learning Techniques," *Proc. 2019 3rd IEEE Int. Conf. Electr. Comput. Commun. Technol. ICECCT 2019*, pp. 1–6, 2019, doi: 10.1109/ICECCT.2019.8868979.