Hybrid Product Recommendation System Using Machine Learning

Shreya Kamath , Uma N

Computer Science Department, New Horizon College of Engineering
Outer Ring Rd, near Marathalli, Kaverappa Layout, Kadubeesanahalli, Kadabeesanahalli, Bengaluru, Karnataka 560103

Abstract— **Hybrid recommendation system is reputed to reduce all the problems and issues that arises alongside FP growth algorithm, hybrid approach involves various algorithms that are part of the solution, by combining algorithms together the machine learning algorithms will be more accurate and the drawbacks will be reduced..**

Keywords— **Algorithms, logistic regression, FP growth, Evolutionary programming, Machine learning.**

I. INTRODUCTION

It is a known fact that all ecommerce website have recommendation system first introduced by Amazon and its recommendation system, ever since Amazon has bought out its own recommendation Systems on its online website every website which is on ecommerce platform also wants a system of their own, what exactly does a product recommendation system do it is it does exactly what it is named after when a customer buys a product the seller urges the user the customer to buy something else alongside, or to at least look at the thing beside the item that they are interested to buy, this is easier when it is in person purchasing because you can the seller can visible see what the year is buying. But virus it becomes a problem when it is online where there is no information or background knowledge on the seller and the customer the seller has no clue who the customer is what are interested and so the instead of calculating the relevant data it is more efficient to calculate the data that is only available and which I which is necessary for the computation of a recommendation.

Recommendations can be calculated by using frequency counting algorithms, such a good algorithm is FP growth. This algorithm is infamous for its branching structure and how much expensive it can be, for every reading of data it can grow really large and can take more time, the drawback of FP growth algorithm can be reduced by using a hybridized approach. Exactly how will hybridized approach solve its problem is somewhat complicated to explain. The project itself deals with the accuracy of recommendation and nothing else matters except for the accuracy of the final decision that is taken by the machine learning algorithm. The algorithms are given input from various online data sets and data bases such a real time data sets which is very valuables kraggle which is in famous for its abundance of data and community.

When a customer comes online to assume what the customer will buy in the first hand itself is a confusing category on top of that it will be easy if a customer would buy an item so we could recommend something based on the recent purchase the user has made.

This becomes a problem because when a purchase done there are a number of items that are available in the shops so therefore we will have 2 the power of n problem, suggesting an item to the user. There are of course other problems associated with the same category of recommendation such as Cold start problem, synonym problem, scalability problem and grey sheep problem. Which are the various problems that are faced by a take a typical recommendation system that is built on machine learning for a base so how does a sucker simple when people do not make any purchases or if they are new to the website then this problem occurs because the system doesn't know what a comment the user, this becomes a long-term problem machine learning module itself because it is unable to make accurate decisions because of lack of previous knowledge of the user. To solve this kind of approach we would have to amalgamate various types of machine learning algorithms that come under supervised learning, it cannot be done using unsupervised learning as we are not sure of the product outcome itself.

II. SUBJECT IN DISCUSSION IN THE PAPER

This problem can be avoided when we use more than one type of algorithm with their own positive and negative pros and cons together alongside FP growth algorithm, for example in understand that there is a particular algorithm that is logistic regression it is good at analyzing small set of data but when it comes to last set of data IT sales this is not a problem for FP growth algorithm so what we could do is record accordingly see the advantages and disadvantages of all these algorithms that we are using in our hybridized approach and segregate them based on their purposes. The ones which have drawbacks can be overcome by using another algorithm. To first letters distinguish our procedure of application into two categories first will be a filter one and the second one will be a filter to filter one will be containing the model based collaborative filtering, such as naive Bayes and Logistic regression. Following this will be the filter to which contains cosine similarity and KNN algorithm. It is initially planned

out so that such that reflected one will be preceding the filter to and the output of filter to will be given to filter 1 therefore the algorithms in the second filter need not have to do the same calculation that is done efficiently much more efficiently than in the algorithms in the filter to repeatedly.

First we will understand how this procedure is work when it is also put together will individually the algorithms function pretty well with their pros and cons all the when they are put together the whole integration testing is very important when whole of the individual machine learning algorithms are put together they need to function together as a unit and also maximize the efficiency and accuracy of the prediction because at the end of the day product recommendation, is also a type of prediction. Whether we can use for the analysis and its well for machine learning algorithms nice base basically works on the probability of something that has already happened it requires the previous even to do a prediction therefore we will not be using that first instead will go with Logistic regression and Logistic regression on the other hand just is very constructive and predicts interaction across the other products with one product in mind this is comparatively faster in terms of working you can also implement it alongside a confusion Matrix for better accuracy all the nice base is much quick and not independent of the data it can solve multiclass predictions therefore this can be used following the Logistic regression section.

Following this we also have to take into consideration the monthly weekly and also weekends most people are viable for shopping during weekends because they are free and their time on their hands where is on weekdays it is blessed to be hard to understand that if there is an offer and if it's a week and then there's the beautiful highest chance of the person buying things online use problem faced by retailers even online and offline is that people do a lot of Window shopping this is a concept where people on the view the items or least put it in their card and do not bother to buy the product this is a loss for the seller because there is no point of a consumer just looking at an item and appreciating it from afar number of sales is the only thing that the seller is concerned about and also the customer satisfaction after the sales process so how can we proceed with this part will the recommendation system that is proposed here can be done using a monthly or a seasonal approach , this way where users' needs and wants can be satisfied in particular let us assume for example sweaters are something that on this winter season and people wear sweaters in the wind, they are exclusively made and what during winter Seasons especially during the cold to cold is parts of winter's so what can we analyse from this is the recommendation system will recommend sweaters to somebody during the winter then of course there is a higher chance of the person buying it but if the person on the recommendation system is recommending sweaters to a user during the season of Summer there is a good chance that the user will never buy it because it is a warm season so that doesn't mean that a user hates the product or something we

had to also consider the users and dislikes to pinpoint the user's likes so basically the user will not dislike sweaters but not at that particular point of time that is not what the user needs at the time of Summer sweaters is something that user will never use during summers so instead of rejecting the recommendation you could actually postponed due to a particular season so only recommend the sweater during winter or you could recommend it again to the user during winter.

In the face of Logistic regression we have the steps for any data pre-processing such as analysing the data for staff will be collecting the data, analysing the data, applying the data by the means of data wrangling and training and testing the data and finally last but not least the accuracy check we can use data frames in machine learning for this because it is a two dimensional data and followed by the first filter we have the second filter that focuses on model based collaborative approach which comes which can which contains cosine similarity and KNN approach following this we have the cosine similarity is a correlation based method can actually find out how similar one is to the other one item is how much similar to the second one followed by that we can use this to analyse various websites this method has been used in family for various online websites where you can track down particular words which are repeating is several times and we can actually figure out that What is the chance of the user Google in the same topic next, we can take an article that has the word cream or mango or not so there's a good chance that the user will Google cream based or some type of condiment relating to mango.

By this we can calculate its similarity with the other items so if you use off by seven item we can make a similarity with another item which will be highly recommended for the next purchase, it is not easy it is not an easy task to make recommendations to a user considering that the user has or has not been exposed to particular environment we will be using Logistic regression for understanding the relationship between the entities the discrete and categorical data is good for calculation for multiclass analysis we will also be considering if it's that the region of the user in specific if the user is it really useful in that way if the user is a high income this will be very beneficial if implemented.

III. ANALYSIS TOOLS IN MACHINE LEARNING FOR BIOINFORMATICS

Below are the list of available online tools and categories that we could use for our analysis.

A. *Machine learning Supervised Algorithms*

When Machine learning algorithms can be used to add this sort of calculations without actually knowing or without going into detail of the mathematical and statistical formula is most of these functions are already available with module what we could do is

we could create a simple structure that takes Input and followers does processing and takes output

B. Supervised Machine Learning methods

- We are using supervised machine learning supervised machine learning is basically the method where it is a sub category of learning it is used for the day labelled data set anything that is a labelled can be used for supervised analysis, Under supervised learning we have to type that is regression and classification regression can be used for estimating life expectancy, marketing problems, weather forecasting, and other applications including growth and prediction where is classification is used for retention image classification and identity of a fraud detection this isn't comes under supervised learning methodology.

- Why we should use supervised machine learning in this case is because we will be using regression and classification to classify whether the product is useful for the user or not if the user is going to make an attempt to buy them or not otherwise there isn't a point of classifying the problem itself.

C. Hybrid methodology

How we will be going about the application is we will utilize the various algorithms that are present in already machine learning directories that can be implemented using the various libraries and packages that are available on Python in particular this approach the advantages are several it can be plugged in an API to an already existing website or any ecommerce website that requires this as an add on every e Commerce website online one or maximum to machine learning algorithms and hybrid approach is unique because this not only makes the things or products that her like to buy the user but also it focuses on the things that a user does not like that these things are not recommended to the user, sometimes the things that are produced to the user can be ambiguous one of the major key factor is that in this approach we try to make recommendations on a monthly basis, or seasonal basis some people college where basically every once in two to three months there is a recommendation list predicted for one particular user and their multi user multiple user the product the product that are recommended for that particular season it can be either a winter summer or fall season and this ensures that the products are up-to-date with the user is living in, it is very beneficial in particularly the drawbacks of most and all algorithms are used are nullified to an extent and sense the products on the Internet especially the ones that are sold are not a mystery it is a known fact that 56% of all products sold online are clothes and following that is a small section of gadgets such as phones earphones and electronic items followed by this we have another section that covers home utilities and a few extra products like cosmetics

D. Links and Bookmarks

- https://www.guru99.com/machine-learning-tutorial.html/
- https://towardsdatascience.com/ml-powered-product-categorization-for-smart-shopping-options- 8f10d78e3294
- https://www.ijert.org/research/a-machine-learning-algorithm-for-product-classification-based-on-unstructured-text-description-IJERTV7IS060214.pdf
- https://medium.com/swlh/8-algorithms-to-build-machine-learning-products-36695241682f
- https://www.interaction-design.org/literature/article/product-service-hybrids- when-products-and-services-become-one

E. References

There are two potential problems with the recommender Systems. One is the scalability, which is how quickly a recommender system can generate recommendation [1]. Content-based recommender systems [1] recommend items based on the textual information of an item. a naive hybrid approach for generating recommendation11, the personality diagnosis algorithm [1]. YouTube video recommendation system delivers personalized sets of videos to signed-in users based on their past activity on the YouTube site(while recommendations are also available in a limited form to sign out users)[2]. problem of frequent item set mining is still an interesting and challenging problem in terms of its computational complexity, frequent item set forms the basis of association rule mining and correlation mining and is widely used in market basket analysis[6].Gathering content data about the item (For example- title, author, cost etc. for the books are some of the common content information[2]. discovery method based on FP-growth algorithm, and ultimately realizes professional courses recommendation[3]. Recommendation engine Agent is the core of the recommendation system. This paper designed it as a Multi-Agent system composed by search engine Agent[4]. Personalized recommendation technology is the key and critical technology of e-commerce recommendation system[4], The analysis of experimental results illustrated that the accuracy of hybrid collaborative filtering recommendation is the highest[5].

IV. CONCLUSIONS

So in conclusion there are various algorithms that are good at certain things and they are at certain things it is not necessary that every algorithm should be perfect or one algorithm can be perfect on itself that is not possible because the user needs likes and dislikes changed change over a. of time requirement best thing to do is to make a recommendation system that is very accurate how can it be accurate is simple it has to think

like to user which in this case for understanding we could come by and algorithms not only in random way but instead in a specific order so that we could maximize the accuracy and minimize the drawbacks that is very important it plays a key role in this particular subject, when algorithms are used together we can do a method called stacking or and sampling these two methods enable us to use more than one type of machine learning algorithm together for the same type of procedure or for the same generation of common output it is important and it helps us to cancel out the drawbacks of the previous algorithms so we could utilize only the best of the abilities of individual algorithm and hand pic or select the ones that suit our requirements the most.

So we will have to reduce the sensitivity to data so that the recommendations wouldn't go wrong we would also have to understand that the application should be simple and implement it in a very simple way so the more Complex a project or a algorithm gets the more hard it is to decode although there are exceptions in a few cases and we would also have to focus on needing less training of the data and handling both continuous and discrete data should be dumb in a proper approach, this would be very important because this would make the system highly scalable which is common every database and every product recommendation system that is ever used on website that actually have real consumers and products the recommendation system would have to face multiple problems in particular a lot of data to handle which was very difficult because when the data increases in size on

the scalability increases it should be slower for the algorithm to run and make predictions. Using a hybrid approach we can also avoid cold start problem and scalability issue can also be minimized to the maximum.

REFERENCES

[1] A Scalable, Accurate Hybrid Recommender System Mustansar Ali Ghazanfar and Adam Prugel-Bennett..

[2] Trends, Problems And Solutions of Recommender System Dr. Sarika Jain, Anjali Grover, Praveen Singh Thakur, Sourabh Kumar Choudhary.

[3] Research on FP-Growth algorithm for agricultural major courses recommendation in China Open University system Fang Liu1, Yuhua Su2, Ting Wang3, Jinfeng Fu4, Shougang Chen1, Chuanjin Ju1

[4] E-commerce Personalized Recommendation System Based on Multi-Agent GUO Yan-yan LIU Qi-cheng. 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010)

[5]
[6] A Detailed Analysis of the various Frequent Itemset Mining Algorithms, Uma.N, Dr. Prashanth c s r.