A Study of Multi Objective Recommender Systems using OTTO dataset

Anashua Krittika Dastidar
Computer Science and Engineering
PES University
Bangalore, India
anashuakrittika@gmail.com

Anubuthi K
Computer Science and Engineering
PES University
Bangalore, India
anubuthi.kottapalli@gmail.com

Anagha Rao
Computer Science and Engineering
PES University
Bangalore, India
raoanagha3105@gmail.com

***Abstract*—Recommender Systems have been well developed over the past few years to assist in user decision making but most recommender systems aim to generate most similar items to the user as he or she has searched for i.e accuracy . But in fact recommender systems must not only do that but also include some degree of Novelty and Serendipity which aims to increase the taste of the users and introduces the user to new content which increases user engagement among other factors and allows the company selling the product to showcase more of its items. This article is aimed at summarising the research done in this domain.**

***Index Terms*—Novelty, serendipity , accuracy**

I. Introduction

Recommender systems are usually judged on three metrics.

- 1. Accuracy (i.e How accurately a system can identify the items the user will buy next)
- 2. Novelty (i.e How many new items the system can introduce the user to)
- 3.Serendipity (i.e randomness in the items recommended to the user just to keep the user engaged).

Accuracy is an important metric as users tend to have similar needs for example a person who buys white rice might also buy brown rice but highly accurate recommendations would lead to a situation in which the recommendations are repetitive say for example recommending a different size packet of the same kind of rice bought by a user previously. This would make the system useless as even though users need some similar things from time to time they need to see diversity also . This is where Novelty and Serendipity come into action by recommending our Rice buying user Tea or Coffee or an array of other beverages we introduce randomness in our system but increase the chances of the user buying something after interacting with more products. But too much randomness can also be detrimental as the user many miss out on useful recommendations based on previous iterations.
Thus we must find a balance between the three metrics by compromising among them. Problems in which we make decisions while trying to compromise among various objectives is called a Multi Objective Problem and many such instances of it exist such as a Financial Planning problem or a Portfolio Optimisation Problem. Thus a recommender system can also be framed as a multi object optimisation problem in which we wish to optimize accuracy , novelty and serendipity to find the best model to recommend items to the user.
The data set we use for this task is the OTTO – Multi-Objective Recommender System Data set from Kaggle [1] the data set consist of data of the format sessionId i.e a unique number which identifies the a sessions in which a number of actions can be performed , each session can consist of the following actions, a click or a add to cart action or place order action on any of the products this will be identified by a product id entry in the articleId field and the value 'click' or 'cart' or 'order' respectively in the type field. These values are ordered by their timestamp which is available in the timestamp field. We wish to predict the top 10 next items which will be picked by a user based on his or her pervious session history

II. Literature Survey

The multi-task recommender systems [2] refer to the recommender system which optimize multiple tasks through a joint learning process. Particularly, there are usually some common or shared representations in multi-task recommender systems, such as the shared latent-factors or embedding layers in the neural network models.
In this category, most of the research utilized Multi Object Evolutionary Algorithms(MOEA) as the Multi Object Optimisation (MOO) algorithm to solve the problems. Take Chai, et al. [3]'s work for example, they built item-based collaborative filtering to produce the top-K recommendation list for each user first. Afterwards, they utilized a multi-objective immune algorithm (MOIA) to learn a top-N recommendation list for each user, while the top-N items will be learned and selected from the top-K items (N¡¡K). The MOIA can learn such top-N recommendation lists by considering accuracy and diversity as the multiple objectives, and finally produce a Pareto set. The authors tried to use PROMETHEE [4] to select the single best solution from the Pareto set.
In B. Geng, et al[6] the main goal was to improve the diversity of the recommendation list without reducing the accuracy and to achieve this a framework which combines collaborative filtering technology with an MOEA was used. Here two functions one which represents the accuracy of the system while another which represents the diversity i.e Novelty and Serendipity in the system are used and the task of creating the

recommendation list is treated as a multi object optimisation problem. First , collaborative filtering generate the candidate solutions, then they use multi-objective algorithm to maximize the matching function and diversity function at the same time. The outputs of the proposed method are a series of Pareto solutions corresponding to the recommendation lists for the target user.

In recommender systems which generate top K potential items for each user usually have the following steps the first step is candidate generation in which we get a set of pareto solutions and the second step is ranking of the generated sets . This second step in e-commerce systems has a number of issues for one of which is multi objective ranking here Click-Trough Rate (CTR) and Conversion Rate (CVR) are two important objectives a highly accurate algorithm will have very high CTR as will predict with very high accuracy the items the user will click on next but will have bad CVR i.e the items bought will not be ordered as this stage is highly dependent on user behaviour and thus the model must generalise well to have high CVR , Yulong Gu, et al.[7] propose a Deep Multifaceted Transformers(DMT), which exploits multiple transformers to model users diverse behavior sequences, utilizes Multi-gate Mixture-of-Experts[8,9] to jointly optimize multi-objectives, and uses a Bias Deep Neural Network for reducing the bias in E-commerce Recommender Systems. This is an improvement over the previously existing Multi-gate Mixture-of-Experts[8,9] models for the same problem as it is crucial to formulate users' different interests based on multiple types of behaviors for significant improvement in multiple objectives. Furthermore, existing works usually ignore the bias issue when they model users' behaviors, which is extremely important in real-world systems.

The work by Ribeiro, et al. [5] provided another way to utilize MOEA in recommender systems by considering accuracy, novelty and diversity. More specifically, they built multiple rating prediction recommendation models, and aggregated the predicted ratings by a linear weighted sum. The aggregated rating can be used to produce the top-N recommendation list so that they can measure accuracy, diversity and novelty

The goal of the research by Karabadji, et al. [10] was to improve accuracy and diversity in the recommendations. By contrast, they proposed an approach which was specifically designed for user-based collaborative filtering (UBCF). The underlying assumption in their work is that the diversity of the recommendations in UBCF is dependent with the diversity of the user neighborhood. As a result, they proposed their approach in which they select the user neighborhood by considering two objectives – the similarity between the target user and the user neighbors, and the dissimilarity among the user neighbors (i.e., the intra-group diversity in the neighborhood). They utilized a weighted sum as the scalarization method to aggregated these two objectives, and adopted genetic algorithm to learn the user neighborhood, as the MOO process. One of the novel methods to solve the problem at hand has been proposed by Li, et al[12], this paper models the process of a recommender system as a multiobjective optimization problem, a multiobjective discrete particle swarm optimization algorithm is proposed to solve the modeled optimization problem.

III. Data Pre-Processing

A. Removing data

The sessions which had less than 10 clicks , less than 5 additions to cart , less than 2 orders were removed . These sessions tend to be very less in number and have the very less activity as the they donot contribute anything essential to the experiments they are removed.

B. Exploratory Data Analysis

OTTO is a large dataset of e-commerce sessions. Sessions consist of user events such as clicking to a product, adding a product to the cart, or ordering a product. Each session belongs to a unique user and the duration of the sessions are different from each other.

- About the Dataset:(Fig 1)
- session: unique ID of the user session
- aid: unique ID of the product
- ts: timestamp of the event
- type: category of the event i.e., whether a product was clicked, added to the user's cart, or ordered during the session

```
session_type,labels
42_clicks,0 1 2 3 4 5 6 7 8 9 10
42_carts,0 1 2 3 4 5 6 7 8 9 10
42_orders,0 1 2 3 4 5 6 7 8 9 10
```

Fig. 1. Structure of the data

In this projects, a "session" actually means a unique "user". So we have to predict what each of the 1,671,803 test "users" (i.e. "sessions") will do in the future. For each test "user" (or "session") we must predict what they will click, cart, and order during the remainder of the week-long test period.

- 216.7M events in training set and 6.9M events in test set
- 12.8M sessions in training set and 1.6M sessions in test set
- 1.8M products in training set and 783K products in test set
- 194.7M clicks in training set and 6.2M clicks in test set
- 16.8M carts in training set and 570K carts in test set
- 5M orders in training set and 65K orders in test set

The sessions in the test set are much shorter than that in the training set. Training and test set are split by time. Training set is 4 weeks of user events and test set is the following week after training set. Proportions of event types are similar in training and test set which can be seen from the visualization. Although the no of clicks , cart additions

and orders are vastly different which is to be expected
As we already know, there are 12.8M unique users (sessions)

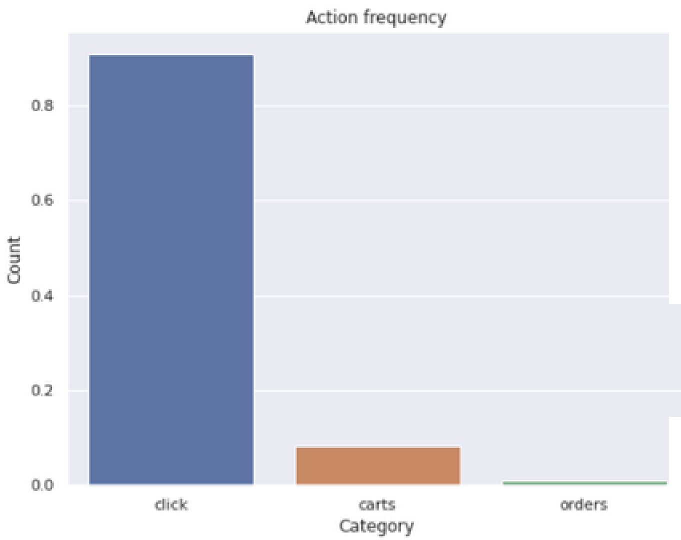


Fig. 2. Training Data

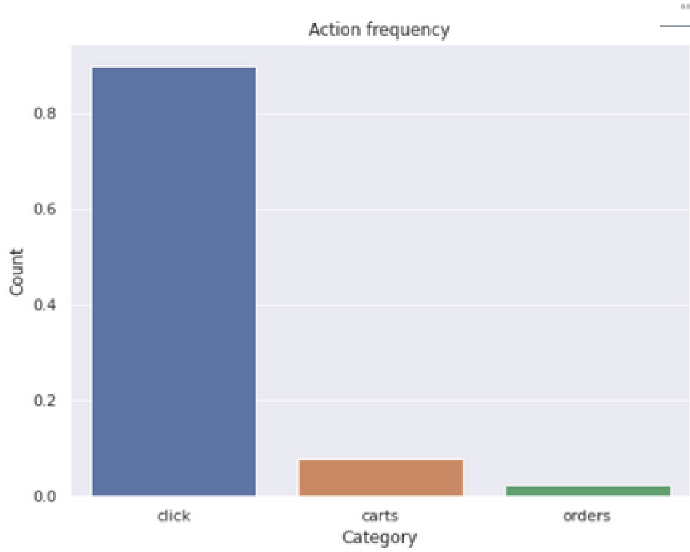


Fig. 3. Testing Data

in training set and 1.6m unique sessions in test set. However, their statistics are also different between training and test set. In training set, mean session duration(Number of actions in 1 session) is 4 times longer than test set. Shortest session in training set is 2 because it wouldn't be possible to create ground-truth on single event sessions. Longest session in training set is 500 which looks like a threshold for trimming extremely long sessions.
Discrepancy in session statistics may not be a problem because training set is 4 weeks while test set is only a single week. Mean session duration being 4 times longer in training set is also an artifact of that. Distributions of session duration can be seen from the visualization Fig 4 and Fig 5. Both densities are on log scale for interpret-ability.

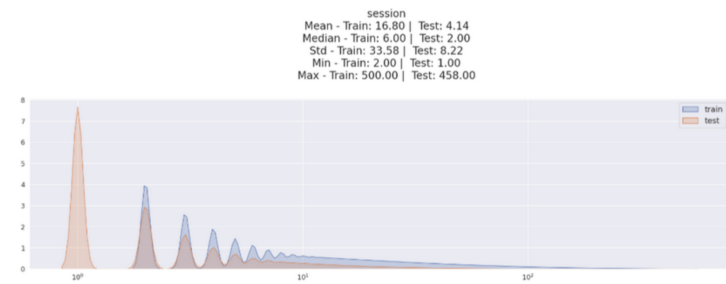


Fig. 4. Session

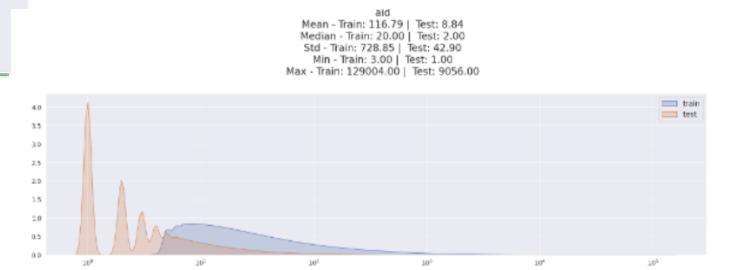
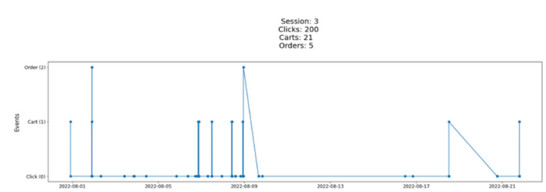
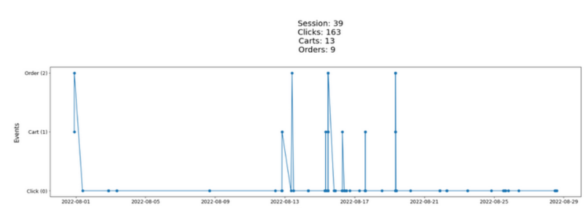
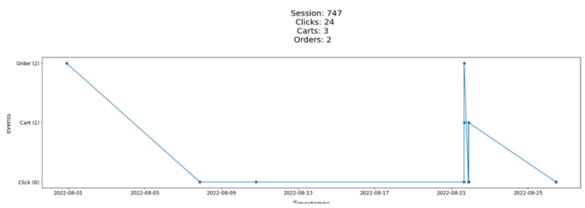


Fig. 5. Aid

Aid frequencies in training and test are very similar. There are minor differences between them since test set is almost 20 times smaller than training set. Computing aid frequencies on concatenated training and test set would be more reliable if aid frequencies are utilized in any way.
Another important thing to consider is how sessions start and end. There are two interesting things here. First, sessions are supposed to start with clicks but very few of them start with carts or orders. Those sessions are less than 1 percent and they might be truncated from their beginning so they could fit into the selected time frame.







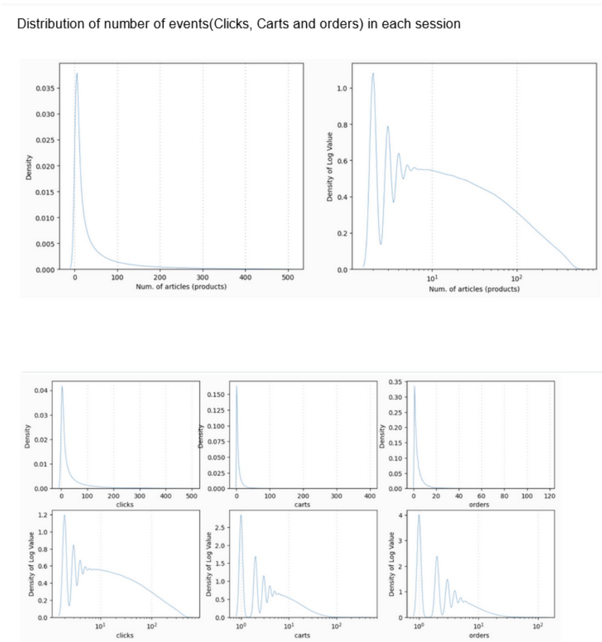


Fig. 6.

IV. Our Work

A. Using Word2Vec

In the paper Miklolov, et al[11] a word embedding architecture is described where we can try to predict the current word in a sentence based on the context(i.e the words before and after it) which is known as the Continuous Bag Of Words model(CBOW) or we can try to maximise the classification of a word based on the other words in the same sentence (continuous Skip-Gram model). In the continuous skip gram model each current word is used as an input to a log-linear classifier with continuous projection layer, and predict words within a certain range before and after the current word. Our Task is modeled as a sequence of articleIds i.e the unique ids representing products along with the type of the task done i.e a click or a cart addition or a order this is fed into the model sequence by sequence . Then for each session on our test data set we predict the next k top article id , action pairs.

In usual multi object recommender systems tasks the idea is to split the system into two parts the first is the one which predicts with very high accuracy the next items to be picked and the second is which introduces random noise to the prediction to add Novelty and Serendipity. Thus for our final predictions we pick the top 10 articles from the recommended list and also pick 10 more randomised values from the k-10 remaining predictions to add to Novelty and serendipity. We believe this process can be further improved and will be exploring such things in the future scope.

Conclusion

In our study of multi object recommender systems we found the work of Chai, et al. [3] and B. Geng, et al[6] extremely interesting and wished to implement a similar two step model but unlike the former which uses collaborative filtering our data set is way too large for us to store such a matrix and thus we wished to improve upon it by finding alternatives for the first step. Some possible solutions could be providing a approach such as implemented in Yulong Gu, et al.[7] as they try to model user behaviours which is very useful when we have abundantly available session data and also this method is efficient as it uses transformers which can be optimised to run efficiently on GPU's. In the end for the second step of the process instead of doing a randomised selection we wish to use evolutionary algorithms to find an Pareto optimal set as it has lead to much better results in similar research problems. We wish to explore these ideas in the future scope of this article.

Acknowledgment

We would like to thank our teachers for the Data Analytics course at PES University for presenting us with the opportunity to study and learn more about the intricate problems of recommender systems. We would like to thank them for their continuous support and help in understanding these concepts.

References

[1] Philipp Normann, Sophie Baumeister, Timo Wilm, "OTTO Recommender Systems Dataset: A real-world dataset of anonymized e-commerce sessions for multi-objective recommendation research"2022-11-01.

[2] Ning, X., Karypis, G. (2010, October). "Multi-task learning for recommender system. In Proceedings of 2nd Asian Conference on Machine Learning "(pp. 269-284). JMLR Workshop and Conference Proceedings.

[3] Chai, Z., Li, Y., Zhu, S. (2021)." P-MOIA-RS: a multi-objective optimization and decision-making algorithm for recommendation systems."Journal of Ambient Intelligence and Humanized Computing, 12, 443-454.

[4] Wikipiedia page for Preferenceranking Organization Method for Enrichment Evaluation

[5] Ribeiro, M. T., Lacerda, A., Veloso, A., Ziviani, N. (2012). Pareto-efficient hybridization for multi-objective recommender systems. In ACM RecSys 2012.

[6] B. Geng, et al,"NNIA-RS: A multi-objective optimisation based recommender system," Physica A(2015), http://dx.doi.org/10.1016/j.physa.2015.01.007

[7] Yulong Gu1, Zhuoye Ding1, Shuaiqiang Wang2, Lixin Zou3, Yiding Liu2, Dawei Yin2 "Deep Multifaceted Transformers for Multi-objective Ranking in Large-Scale E-commerce Recommender Systems" CIKM '20, October 19–23, 2020

[8] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H Chi. 2018. Modeling task relationships in multi-task learning with multi-gate mixture-ofexperts. In KDD'18. 1930–1939.

[9] Xiao Ma, Liqin Zhao, Guan Huang, Zhi Wang, Zelin Hu, Xiaoqiang Zhu, and Kun Gai. 2018. Entire space multi-task model: An effective approach for estimating post-click conversion rate. In SIGIR'18. 1137–1140

[10] Karabadji, N. E. I., Beldjoudi, et al. (2018). Improving memory-based user collaborative filtering with evolutionary multi-objective optimization. Expert Systems with Applications, 98, 153-165.

[11] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean Efficient Estimation of Word Representations in Vector Space,7 Sep 2013

[12] Zhaoxing Li, Chen Li, Zhang Jue Multi-objective Particle Swarm Optimization Algorithm for Recommender System