

**NEURAL COLLABORATIVE FILTERING BASED
RECOMMENDATION SYSTEM FOR
PURCHASED PRODUCT RECOMMENDATION**

Dushyantha Widanagamage

209395L

MSc in Computer Science Specialising in Data Science Engineering and
Analytics

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa

Sri Lanka

November 2022

**NEURAL COLLABORATIVE FILTERING BASED
RECOMMENDATION SYSTEM FOR
PURCHASED PRODUCT RECOMMENDATION**

Dushyantha Widanagamage

209395L

Thesis submitted in partial fulfilment of the requirements for the degree
MSc in Computer Science Specialising in Data Science Engineering and
Analytics

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

November 2022

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date:

The above candidate has carried out research for the Masters Dissertation under my supervision.

Name of the supervisor:

Signature of the supervisor:

Date:

ACKNOWLEDGEMENTS

Throughout the writing of this dissertation I have received a great deal of support and assistance. I would first like to thank Dr. Sapumal Ahangama, for the invaluable help, guidance and feedback provided to me throughout this research. His expertise and guidance was the key to complete this research project successfully.

Furthermore, I would like to express my gratitude to my friends, who had helped me to find invaluable resources and sharing their invaluable thoughts. I also, like to express my gratitude to my Parents, for the loving and kind help they provided and the encouragements given throughout my life.

ABSTRACT

In order to validate that the problem exists, I followed the procedure as explained below.

First I grouped the data set with user id and the product. Then, for each user and item, I have derived the number of views, transactions and add to cart events. Then, I have created 10 new data sets. For the first five data sets, I have assigned different weights based on the event type (i.e. view, purchase or transaction). As for the second five data sets, they were created with different volumes of view, transaction and purchased events. Then I have verified that, with the presence of outliers (view events), the purchased products are not recommended to the user. To verify this behaviour I have used Bayesian Personalized Ranking, Neural Collaborative Filtering, Generalized Matrix Factorization, Most Pop, Item KNN adjusted and Multi-Layer Perceptron models.

Thereafter, I have removed view data from the data set and grouped data records based on the product and user. Next I have used a weighting scheme combined with binning to derive a rating score.

Next, I have used four models to verify my solution. These includes, Bayesian Personalized Ranking, Neural Collaborative Filtering, Item KNN adjusted, Generalized Matrix Factorization and Multi-Layer Perceptron. I have used fivefold cross validation to train the models and used a separate data set for validation. The results were promising. I received a Hit ratio 0.275 for HR@10. This was a major improvement as, before this the Hit ratio was near to 0.

TABLE OF CONTENTS

Declaration	ii
Acknowledgement	iii
Abstract	iv
Table of content	v
List of Figures	vii
List of Tables	xi
List of abbreviations	xiii
1 Introduction	1
1.1 Reasearch Problem	2
1.2 Objectives	4
1.3 Reasearch Contribution	5
2 Litrature Review	8
2.1 Training and Test Data Splits	9
2.2 Feature Engineering in Recommendation Systems	10
2.3 Collaborative Filtering	11
2.3.1 Matrix Factorization Algorithm	11
2.4 Content Based Recommender Systems	21
2.5 Hybrid Recommender Systems	26
2.6 Outlier Detection	27
2.7 Evaluation Methods	30
2.8 Hyper Parameter Optimization	31
2.9 Deploying Recommendation Systems	32
2.10 What is Bias in Recommendation Systems?	33
2.10.1 Types of Bias and Mitigation Techniques	34
3 Methodology	38
3.1 Problem Verification	38
3.2 Experiment Detail And Solution	41
4 Results	46
4.1 Problem Verification Discussion	46
4.2 Effect of Learning Rate on Model Performance	60

4.3	Effect of Predictive Factor size on Model Performance	61
4.4	Top K recommendation and model performance discussion	63
4.5	Effect of Pretraining on model performance discussion	65
4.6	Effect of change of MLP layers on model performance	67
4.7	Comparison with the existing research	68
4.7.1	Comparison of bench mark data set with Retail Rocket Data Set	68
4.7.2	Comparison of Preprocessing techniques used in existing research and Retail Rocket Data Set	78
4.7.3	Comparison of feature engineering techniques used in existing research and Retail Rocket Data Set.	83
4.7.4	Training Procedures used in Research and one used in the research	84
4.7.5	Compare different results obtained in research with this research	90
5	Related Work	94
6	Conclusion and Recommendation	95
7	References	98

LIST OF FIGURES

Figure 2.1: Objective function for Matrix Factorization	12
Figure 2.2: Frobenius distance between two matrixes	12
Figure 2.3: Weighted Objective function used in recommendation systems	12
Figure 2.4: User interface of the learning system	33
Figure 4.1: BPR - NDCG@10 and HR@10 performance for different weights and volumes for view, purchase and addToCart events.	46
Figure 4.2: NCF NDCG@10 and HR@10 performance with different weights and volumes for view, purchase and addToCart events.	47
Figure 4.3: GMF - NDCG@10 and HR@10 performance for different weights and volumes for view, purchase and addToCart events.	47
Figure 4.4: MLP - NDCG@10 and HR@10 performance for different weights and volumes for view, purchase and addToCart events.	48
Figure 4.5: Most Pop - NDCG@10 and HR@10 performance with different volumes and weights of view, purchase and addToCart events.	49
Figure 4.6: ItemKNN NDCG@10 and HR@10 performance with different volumes and weights of view, purchase and addToCart events.	49
Figure 4.7: Comparison of NCF model output score distribution with training score distribution for the data set 9. Left training score distribution and right model output score distribution.	51
Figure 4.8: Comparison of MostPop model output score distribution with training score distribution for the data set 9. Left training score distribution and right model output score distribution.	52

Figure 4.9: Comparison of ItemKNN model output score distribution with training score distribution for the data set 9. Left training score distribution and right model output score distribution.	52
Figure 4.10: Comparison of NCF model output score distribution with training score distribution for the data set 10. Left training score distribution and right model output score distribution.	53
Figure 4.11: Comparison of MostPop model output score distribution with training score distribution for the data set 10. Left training score distribution and right model output score distribution.	53
Figure 4.12: Comparison of ItemKNN model output score distribution with training score distribution for the data set 10. Left training score distribution and right model output score distribution.	54
Figure 4.13: Comparison of NCF model output score distribution with training score distribution for the data set 7. Left training score distribution and right model output score distribution.	54
Figure 4.14: Comparison of MostPop model output score distribution with training score distribution for the data set 7. Left training score distribution and right model output score distribution.	55
Figure 4.15: Comparison of ItemKNN model output score distribution with training score distribution for the data set 7. Left training score distribution and right model output score distribution.	55
Figure 4.16: Comparison of NCF model output score distribution with training score distribution for the data set 2. Left training score distribution and right model output score distribution.	56
Figure 4.17: Comparison of Most Pop model output score distribution with training score distribution for the data set 2. Left training score distribution and right model output score distribution.	56

Figure 4.18: Comparison of Item KNN model output score distribution with training score distribution for the data set 2. Left training score distribution and right model output score distribution.	57
Figure 4.19: Comparison of NCF model output score distribution with training score distribution for the data set 4. Left training score distribution and right model output score distribution.	57
Figure 4.20: Comparison of Most Pop model output score distribution with training score distribution for the data set 4. Left training score distribution and right model output score distribution.	58
Figure 4.21: Comparison of ItemKNN model output score distribution with training score distribution for the data set 4. Left training score distribution and right model output score distribution.	58
Figure 4.22: Learning rates and model NDCG@10 performance.	60
Figure 4.23: Learning rates and model HR@10 performance.	60
Figure 4.24: Embedding size and model NDCG@10 performance.	61
Figure 4.25: Embedding size and model HR@10 performance.	62
Figure 4.26: Top k and model NDCG@10 performance.	63
Figure 4.27: Top k and model HR@10 performance.	64
Figure 4.28: Pre trained models and model NDCG@10 performance with embedding size.	65
Figure 4.29: Without pre training models and model NDCG@10 performance with embedding size.	66
Figure 4.30: Increase MLP layers, and embedding size with model NDCG@10 performance.	67

Figure 4.31: Increase MLP layers, and embedding size with model HR@10 performance.	68
Figure 4.32: Clustering for the data set with item, visitor and event counts (Heat Map)	74
Figure 4.33: Cluster tree created for the data set with item ,visitors and event counts	75
Figure 4.34: Centroid Chart for the item id, visitor id and event counts	75
Figure 4.35: Cluster table item id, visitor id and event count	76
Figure 4.36: Scatter plot view for the cluster 0 and 1	76
Figure 4.37: Scatter plot view for the second cluster	76
Figure 4.38: Scatter plot view for the cluster 3	77
Figure 4.39: Bar chart for the visitor distribution	77
Figure 4.40: Item id distribution (Bar Chart)	77
Figure 4.41: Event count distribution (Bar Chart)	78
Figure 6.1: Results for the given criteria	96
Figure 6.2: Recommended products for a given result	97
Figure 6.3: Internal Representation of a knowledge graph	97

LIST OF TABLES

Table 1.1: Data Quality Dimensions Describing Data (Measurements)	3
Table 2.1: Comparison of Stochastic Gradient Descent and Weighted Alternating Least Squares	13
Table 3.1: Events and assigned weights	38
Table 3.2: Events and added data volumes	39
Table 3.3: Parameters used initialize the models	39
Table 3.4: Events and assigned scores	41
Table 3.5: Model Learning rate change and recorded HR@10 and NDCG@10	42
Table 3.6: Model Predictive factor change and recorded HR@10 and NDCG@10	42
Table 3.7: Top K recommendation evaluation results and NDCG (When K changed)	43
Table 3.8: Top K recommendation evaluation results and Hit Rate (When K changed)	43
Table 3.9: Model parameters used to initialize pretraining models in experiment 4	44
Table 3.10: Neural Collaborative Filtering evaluation results when is used Pretraining and Not and reported HR@10 and NDCG@10	44
Table 3.11: Neural Collaborative Filtering evaluation results and MLP Layers. Models evaluated in NDCG@10	45
Table 3.12: Neural Collaborative Filtering evaluation results and MLP Layers. Models evaluated in HR@10	45
Table 4.1: Existing Benchmark Data Set Statistic Comparison	68

Table 4.2: Compare preprocessing techniques applied in different research	78
Table 4.3: Compare different feature engineering techniques applied in research	83
Table 4.4: Compare different training and evaluation procedures used in research	85
Table 4.5: Compare different results obtained in different research	91

LIST OF ABBREVIATIONS

Abbreviation Description

Pop	Item popularity based, non-personalized baseline method.
LSTM	Long short-term memory
RNN	Recurrent Neural Networks.
CF	Collaborative Filtering
OCCF	One Class Collaborative Filtering
RMSE	Root Mean Square Error
NDCG	Normalized Discounted Cumulative Gain
MPR	Mean Reciprocal Rank
HR	Hit Ratio
Caser	Convolutional Sequence Embedding Recommendation Model
CNN	Convolutional Neural Networks
AUC	Area under curve
ALS	Alternating Least Square
ARP	Average Recommendation Popularity
APLT	Average Percentage of Long Tail Items
ARP	Average Recommendation Popularity
ACLT	Average Coverage of Long Tail items
MF	Matrix Factorization