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

LIST OF FIGURES

LIST OF TABLES

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