A COLLABORATIVE FILTERING BASED RECOMMENDER SYSTEM ALLEVIATING THE MOST COMMON CHALLENGES

A DISSERTATION

Submitted in Partial fulfilment of the Requirement for the Award of the Degree of

MASTER OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted By:

Sahil Manchanda

2K16/CSE/12

Under the Guidance of:

Mr Manoj Sethi

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahabad Daulatpur, Main Bawana Road, Delhi-110042

JUNE 2018

DECLARATION

I hereby declare that the Major Project-II work entitled "**A Collaborative Filtering based Recommender System Alleviating the most common Challenges**" which is being submitted to **Delhi Technological University**, in partial fulfilment of requirements for the award of degree of Master of Technology (Computer Science and Engineering) is a bona fide report of Major Project-II carried out by me. The material contained in the report has not been submitted to any university or institution for the award of any degree.

Sahil Manchanda

2K16/CSE/12

CERTIFICATE

This is to certify that Project Report entitled "**A Collaborative Filtering based Recommender System Alleviating the most common Challenges**" submitted by **Sahil Manchanda (Roll no. 2k16/CSE/12)** for partial fulfilment of the requirement for the award of degree Master Of Technology (Computer Science and Engineering) is a record of the original work carried out by him under my supervision. To the best of my knowledge, this work has not been submitted in part or full for any degree or Diploma to this University or elsewhere.

Mr Manoj Sethi
Project Guide
Department of Computer Science & Engineering
Delhi Technological University

ACKNOWLEDGEMENT

First of all I would like to thank the Almighty, who has always guided me to work on the right path of the life. My greatest thanks are to my parents who bestowed ability and strength in me to complete this work.

I would like to express profound gratitude to my project guide **Mr Manoj Sethi** who has been a constant source of inspiration to me throughout the period of this project.

I am grateful to Dr. Rajni Jindal, HOD, Computer Science & Engineering Department, DTU for her immense support. I would also like to acknowledge Delhi Technological University library and staff for providing the right academic resources and environment for this work to be carried out.

Sahil Manchanda

2K16/CSE/12

ABSTRACT

Recommender Systems (RS) is a subclass of information filtering system that seeks to predict the "rating" or "preference" that a user would give to an item. These are the Systems for recommending items (e.g. books, movies, web pages, newsgroup messages) to users based on examples of their preferences. With increasing information availability online, it has become extremely challenging to build accurate Recommender Systems. But at the same time, they are extremely important for businesses all around the globe. Hence it becomes vital to study these systems and improve their accuracy.

Recommender Systems are generally classified into Content based and Collaborative filtering(CF) based RS. Content based RS focus on the properties of the items while CF based RS estimates user/item similarity. Here, we discuss RS' with special emphasis on CF based techniques which are further classified as user/item based CF Algorithms.

The Issues and Challenges faced by CF based RS makes it an active area of research with many new variants and improvements being proposed in the past years. This work proposes a recommender system that alleviates three challenges simultaneously i.e. shilling attacks, sparsity problem and cold start problem that are most common in this domain. As a result, the proposed system proves to be more effective as compared to the conventional algorithm. Practically, this system can be utilized to produce quality recommendations to its users, as a result of which, the profitability can be maximized. The dataset used for doing the analysis and building the system is the popular MovieLens dataset. The work also presents a study of the Collaborative Filtering Improvements along with analysis of the important aspect of each improvement. This study can be utilized by any researcher working in this direction to have a complete idea of the past and the latest improvements in this area.

TABLE OF CONTENTS

	Page No.
Declaration	ii
Certificate	iii
Acknowledgement	iv
Abstract	v
Table of Contents	vi
List of Figures	viii
List of Tables	ix
List of Abbreviations	x
1. Introduction	1
1.1 Recommender Systems	1
1.2 Scope of Work	2
1.3 Research Goals	3
1.4 Organization of the thesis	3
2. Literature Survey	5
3. Collaborative Filtering	9
3.1 Overview	9
3.2 Framework of Collaborative Filtering	9
3.3 Approaches to Collaborative Filtering	11
3.3.1 Memory based CF Algorithms	12
3.3.2 Model based CF Algorithms	17
3.3.3 Hybrid Algorithms	18
3.4 Challenges and Attacks to Collaborative Filtering	18
4. Shilling Attacks	20
4.1 Overview	20
4.2 Types of Shilling Attack Models	21
4.3 Principal Component Analysis	22
4.3.1 Eigen Values and Eigen Vectors	23
4.3.2 Steps for Performing PCA	26
4.4 PCA for Shilling Detection	26
4.4.1 Variable Selection using PCA	27
4.4.2 PCA with perturbation	28
5. Alleviating Sparsity and Cold start Problems	30
5.1 Alleviating Sparsity	30
5.1.1 Weighted Slope one Technique	30

5.2 Alleviating Cold start problem 32

6. The Proposed Framework 37
6.1 Architecture of the Proposed System 37
6.1.1 Genre based HELF 38
6.2 Proposed Algorithm 39

7. Implementation and Results 40
7.1 Evaluation Metrics 40
7.2 Implementation Details 41
7.3 Dataset Used 41
7.4 Results 42

8. Conclusion and Future Work 46
8.1 Conclusion 46
8.2 Future Scope 46

References 47

Appendix A: CODE SNIPPETS 51
Appendix B: SNAPSHOTS OF THE SYSTEM 56
Appendix C: PUBLICATION 57

LIST OF FIGURES

Figure No.	Title or Explanation	Page No.
1.1	Various Recommender System Approaches	1
3.1	Framework of Collaborative Filtering	10
3.2	Rating or Utility Matrix	10
3.3	Classification of CF Techniques	12
3.4	Sample Rating Matrix	14
3.5	User Based Pearson Correlation	14
3.6	Item Based Pearson Correlation	15
3.7	User based CF Process	16
4.1	Principal Components obtained by PCA on a sample data	23
4.2	Eigen vectors (shown in red)	24
5.1	Sample rating matrix for weighted slope one scheme	31
5.2	Alleviating user cold start problem using ask-to-rate process	33
6.1	Architecture of the Proposed System	37
7.1	First 5 rows and columns of the dataset	42
7.2	Comparative results without shilling attack	42
7.3	Comparative results with shilling attack	43

LIST OF TABLES

Table No.	Title or Explanation	Page No.
2.1	Comparative Study of Collaborative Filtering research	6
3.1	Comparison of User-based and Item-based approaches	17
3.2	Challenges and Attacks to Collaborative Filtering along with their cause(s)	19
4.1	Attacker's Profile Structure	20
4.2	Features of different types of Attack Models	22
7.1	List of movies using popularity strategy	43
7.2	List of movies using entropy strategy	44
7.3	List of movies using HELF strategy	44
7.4	List of movies using Genre based HELF strategy (Action)	44
7.5	List of movies using Genre based HELF strategy (Adventure)	45

LIST OF ABBREVIATIONS

1. RS: Recommender/Recommendation System
2. CF: Collaborative Filtering
3. ACOS: Adjusted Cosine Similarity
4. PCC: Pearson Correlation Coefficient
5. UBCF: User based Collaborative Filtering
6. IBCF: Item based Collaborative Filtering
7. PCA: Principal Component Analysis
8. WS1: Weighted Slope One
9. HELF: Harmonic Mean of Entropy and Logarithm of Frequency
10. MAE: Mean Absolute Error