PERSONALISED MOVIE RECOMMENDATION BASED ON MULTI MODEL DATA INTEGRATION

J G I Madushanki

199477V

Degree of Master of Science

Department of Computational Mathematics

University of Moratuwa
Sri Lanka

March 2022

PERSONALISED MOVIE RECOMMENDATION BASED ON MULTI MODEL DATA INTEGRATION

J G I Madushanki

199477V

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree Master of Science

Department of Computational Mathematics

University of Moratuwa

Sri Lanka

March 2022

DECLARATION

I declare that this is my own work and this thesis/dissertation2 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 thesis/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:

J. G. I. Madushanki 04th of March 2022

The above candidate has carried out research for the Masters/MPhil/PhD thesis/ dissertation under my supervision.

Signature of the supervisor: Date

04th of March 2022

Abstract

Recommendation systems plays an essential role in the modern era, and it is a part of routine life where it guides the users in a personalised manner towards interesting and useful objects in a large collection of possible options. The aim of the movie recommendation system is to help movie lovers by generating suggestions on what movie to watch. If movie recommender systems are not in place, movie lovers need to spend time on choosing a movie by going through long lists of movies, which is a time consuming task. Therefore, a lot of research has been conducted to generate movie recommendations using different approaches including pure recommendation techniques and hybrid techniques. However, the recommendations generated through these approaches lack personalisation and accuracy.

This thesis presents our approach to generate personalised movie recommendations using multi model data integration to improve the personalisation and accuracy. Different data sources are integrated as inputs when designing this research. A content-based filtering technique collaborated with genetic algorithm-based optimization was utilized for implementation of this research. A precision value of 0.65 was obtained while evaluating the multi-model data integration-based movie recommender system with genetic algorithm-based optimization.

DEDICATION

I dedicate my dissertation work to my family, friends and my supervisor Dr A.T.P. Silva who encouraged me to achieve the final year research project. A special feeling of gratitude to my loving parents whose thoughts and words of encouragement to achieve this. I also dedicate this dissertation to my office mates who have supported me throughout this process. I will always appreciate the kind words and encouragement of them to achieve my goal. I dedicate this work and give many thanks to my supervisor and staff at University of Moratuwa for their help and specially for academic staff for providing guidance throughout this research.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr A.T.P. Silva for providing me guidance and insights throughout this project. This project would not have been possible without the support of my supervisor. Also, I would like to extend my thanks to the University of Moratuwa for providing me the opportunity to carry out this research successfully and for the continuous support provided throughout the research. I would like to pay my sincere gratitude to all the academic and non-academic staff members of University of Moratuwa and the batchmates for their generous support, comments and encouragement throughout the project. Last, but not least, my warm and heartfelt thanks go to my family for their tremendous support and hope they had given to me. Without that hope, this research would not have been possible. Thank you all for the strength you gave me.

TABLE OF CONTENTS

DECLARATION ii
ABSTRACT iii
DEDICATION iv
ACKNOWLEDGEMENTS v
TABLE OF CONTENTS vi
LIST OF FIGURES ix
LIST OF TABLES xi
LIST OF ABBREVIATIONS xii
LIST OF APPENDICES xiii
INTRODUCTION 1
1.1 Prolegomena 1
1.2 Aim and Objectives 1
1.3 Background and Motivation 2
1.4 Problem Definition 4
1.5 Novel Approach to Personalized Movie Recommender System 4
1.6 Resource Requirements 5
1.7 Structure of the Thesis 6
1.8 Summary 6
DEVELOPMENTS AND ISSUES OF PERSONALISED MOVIE RECOMMENDATION SYSTEMS 7
2.1 Introduction 7
2.2 Early Developments of Movie Recommender Systems 7
2.3 Breakthrough in Movie Recommender System 8
2.4 Challenges in Movie Recommender Systems 13
2.5 Problem Definition 18
2.6 Summary 18
TECHNOLOGIES USED FOR MOVIE RECOMMENDER SYSTEMS 19
3.1 Introduction 19
3.2 Recommender Systems 19
3.2.1 Collaborative filtering 19
3.2.2 Content based filtering 20
3.2.3 Hybrid filtering 20

3.2.4 Knowledge based recommender systems 21
3.2.5 Ontology based recommender systems 21
3.2.6 Intelligent recommender system 22
3.2.7 Demographics based recommender system 22
3.2.8 Context aware recommender system 22
3.3 Genetic Algorithm 23
3.3.1 Initial Population 23
3.3.2 Fitness Function 23
3.3.3 Genetic Algorithm Operators 23
3.3.4 Selection 24
3.4 Personalised Movie Recommendation Using Content Based Filtering 24
3.5 Summary 24
APPROACH 26
4.1 Introduction 26
4.2 Hypothesis 26
4.3 Input 26
4.4 Output 26
4.5 Process 27
4.6 Users 28
4.7 Features 28
4.8 Summary 28
DESIGN 29
5.1 Introduction 29
5.2 Pre-Processing Module 29
5.3 Review Analysis Module 29
5.4 Profiling Module 30
5.5 Content Based Recommendation Module 30
5.6 Evolutionary Computing Module 30
5.7 Summary 31
IMPLEMENTATION 32
6.1 Introduction 32
6.2. Implementation 32
6.2.1 preprocessing module 32
6.2.2 Review analysis module 38

6.2.3 The Profiling module 40
6.2.4 Content based recommendation module 43
6.2.5 Evolutionary computing module 44
6.3 Summary 47
EVALUATION 48
7.1 Introduction 48
7.2 Results 48
7.3 Summary 62
CONCLUSION AND FUTURE WORK 63
8.1 Introduction 63
8.9 Summary 63
REFERENCES 64
APPENDIX 68
Appendix A 68

LIST OF FIGURES

Figure 5.1. 1 Top level architecture of personalised movie recommender system 29

Figure 5.6. 1 The main steps involve in genetic algorithm 31

Figure 6.2.1. 1 Mappings in Movie Lens dataset for TMDB Id and IMDB Id 33
Figure 6.2.1. 2 Mappings in Movie Lens dataset for YouTube Id and Movie Id 34
Figure 6.2.1. 3 Preprocessing user data 36
Figure 6.2.1. 4 Java method to get YouTube video details 36
Figure 6.2.1. 5 Python script to get IMDB review data 37
Figure 6.2.1. 6 Python script to create TMDB and IMDB mapping dataset 38

Figure 6.2.2. 1 Naive Bayes classifier to get review sentiment 39
Figure 6.2.2. 2 Integration of review sentiment with movie profiles 40

Figure 6.2.3. 1 Integration of review sentiment with movie profiles 40
Figure 6.2.3. 2 Sample movie profile 41
Figure 6.2.3. 3 Sample review data obtained from IMDB API 41
Figure 6.2.3. 4 Creation of user profiles 42
Figure 6.2.3. 5 Creation of movie profiles 43

Figure 6.2.4. 1 Multi model data integration-based content-based filtering implementation 44

Figure 6.2.5. 1 Architecture of evolutionary computing-based approach 45
Figure 6.2.5. 2 Creation of individuals and chromosomes 46
Figure 6.2.5. 3 Single point crossover implementation 47
Figure 6.2.5. 4 Swap mutation implementation 47

Figure 7.2. 1 Results of recommendation approaches 49
Figure 7.2. 2 Precision obtained for recommendation approaches 50
Figure 7.2. 3 Calculated precision for experiments 50
Figure 7.2. 4 Calculated Degree of novelty per movie 51
Figure 7.2. 5 Calculated degree of novelty for experiments 51
Figure 7.2. 6 Google recommendations for movie Garfield: A Tail of Two Kitties 52
Figure 7.2. 7 Recommended movies to movie Garfield: A Tail of Two Kitties 52
Figure 7.2. 8 Google recommendations for movie Thor 53
Figure 7.2. 9 Recommended movies to movie Thor 53

Figure 7.2. 10 Google recommendations for movie Interstellar 54
Figure 7.2. 11 Recommended movies to movie Interstellar 54
Figure 7.2. 12 Google recommendations for movie Inception..... 55
Figure 7.2. 13 Recommended movies to movie Inception 55
Figure 7.2. 14 Google recommendations for movie The Amazing Spider-Man 2 56
Figure 7.2. 15 Recommended movies to movie The Amazing Spider-Man 2..... 56
Figure 7.2. 16 Google recommendations for movie Toy Story..... 57
Figure 7.2. 17 Recommended movies to Toy story 57
Figure 7.2. 18 Google recommendations for movie Apollo 13..... 58
Figure 7.2. 19 Recommended movies to movie Apollo 13..... 58

LIST OF TABLES

Table 2. 1 Summary of benefits and challenges identified in literature review..... 14

LIST OF ABBREVIATIONS

Abbreviation	Description
CARS	Context aware recommender systems
CSV	Comma Separated Values
DBRS	Demographics based recommender systems
GA	Genetic Algorithm
IRS	Intelligent recommender systems
KBRS	Knowledge based recommender systems
KNN	K Nearest Neighbors
MAE	Mean Absolute Error
OBRS	Ontology based recommender systems
RMSE	Root Mean Square Error
SVD++	Singular value decomposition plus-plus

LIST OF APPENDICES

Appendix - A Questionnaire to identify user interests and behaviors towards movies 69