

Mohammad Amin Rashidifar

# HYBRID MOVIE RECOMMENDERS BASED ON NEURAL NETWORKS AND DECISION TREES



**Anchor Academic Publishing**

*disseminate knowledge*

**Rashidifar, Mohammad Amin: HYBRID MOVIE RECOMMENDERS BASED ON NEURAL NETWORKS AND DECISION TREES, Hamburg, Anchor Academic Publishing 2015**

PDF-eBook-ISBN: 978-3-95489-937-1

Druck/Herstellung: Anchor Academic Publishing, Hamburg, 2015

**Bibliografische Information der Deutschen Nationalbibliothek:**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

**Bibliographical Information of the German National Library:**

The German National Library lists this publication in the German National Bibliography. Detailed bibliographic data can be found at: <http://dnb.d-nb.de>

All rights reserved. This publication may not be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

---

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Bearbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Die Informationen in diesem Werk wurden mit Sorgfalt erarbeitet. Dennoch können Fehler nicht vollständig ausgeschlossen werden und die Diplomica Verlag GmbH, die Autoren oder Übersetzer übernehmen keine juristische Verantwortung oder irgendeine Haftung für evtl. verbliebene fehlerhafte Angaben und deren Folgen.

Alle Rechte vorbehalten

© Anchor Academic Publishing, Imprint der Diplomica Verlag GmbH  
Hermannstal 119k, 22119 Hamburg  
<http://www.diplomica-verlag.de>, Hamburg 2015  
Printed in Germany

# **HYBRID MOVIE RECOMMENDERS BASED ON NEURAL NETWORKS AND DECISION TREES**

---

The internet provides a lot of information to users. To help users find the items of their interest in this information overload, recommender systems have been developed. In this book we explored movie recommender systems based on three recommendation methods: content-based, collaborative filtering and a hybrid recommendation one based on the previous two. The algorithms that we used are the decision tree learning and the neural networks. The algorithms were implemented by using the data mining software Weka. To test these recommender systems, we combined the movie data from the Internet Movie Database and the rating data provided by Netflix. The results show that the proposed hybrid recommender systems does not perform better or worse than the content-based recommender systems and collaborative filtering recommender systems.

# Table of Contents

CHAPTER 1 .....	1
INTRODUCTION .....	1
1.1 Background .....	1
1.2 Motivation .....	2
1.3 Goal .....	2
1.4 Methodology .....	4
1. Study literature.....	4
2. Collect datasets .....	5
3. Implement algorithms .....	5
4. Experiment.....	5
1.5 Structure .....	6
Chapter 2.....	8
Related work.....	8
2.1 Recommender systems.....	8
2.3 Movie recommender systems.....	10
2.4 Neural networks .....	11
2.5 Decision trees learning.....	12
CHAPTER 3 .....	14
NEURAL NETWORKS -BACKPROPAGATION .....	14
3.1 Backpropagation algorithm.....	14
3.2 Neural Networks for content-based recommendation.....	16
3.2.1 Example content-based recommendation.....	17
3.3 Neural Networks for collaborative filtering recommendation .....	20
3.3.1 Example collaborative filtering .....	20
3.4 Neural networks for hybrid recommendation .....	23
3.4.1 Example hybrid recommendation.....	24
CHAPTER 4 .....	27
DECISION TREE LEARNING – C4.5.....	27
4.1 C4.5 algorithm.....	27
4.1.1 Build the decision tree .....	27
4.1.2 Pruning the decision tree .....	31
4.1.3 Classifying cases .....	32
4.2 Decision tree learning for content-based recommendation.....	32
4.2.1 Example content-based recommendation.....	33
4.3 Decision tree learning for collaborative filtering.....	35

4.3.1 Example collaborative filtering .....	36
4.4 Decision tree learning for hybrid recommendation.....	39
4.4.1 Example hybrid recommendation.....	40
4.5 Summary .....	42
CHAPTER 5 .....	44
IMPLEMENTING ALGORITHMS - WEKA.....	44
5.1 Weka.....	44
5.2 Multilayer Perceptron.....	44
CHAPTER 6 .....	50
EXPERIMENTS & RESULTS.....	50
6.1 Datasets .....	50
6.3 Parameter Optimization.....	52
6.3 Results .....	56
6.5 Summary .....	62
CHAPTER 7 .....	64
CONCLUSION.....	64
7.1 Summary .....	64
7.2 Future work .....	67
References.....	69
Appendix A.....	71
Appendix B.....	78
Results of the optimization of the Weka-parameters.....	78

# CHAPTER 1

## INTRODUCTION

### 1.1 Background

Nowadays the World Wide Web provides a new way of communication and has a great impact on both academic research and daily life. A lot of information can be found on the Internet and is easily accessible. In order to help users to deal with the information overload and find the information or items of their interest, so-called recommender systems have been developed. These recommender systems are used for several purposes, like proposing web pages, movies, restaurants, interesting articles and so on. There are various recommendation methods that can be used to find the preferences of a user and each recommendation method has its strengths and weaknesses. To reduce these weaknesses and take advantage of the strengths of different recommendation methods, these methods are combined in hybrid recommender systems. In this book, recommender systems for movies will be examined. The properties, advantages and disadvantages of the movie recommenders and their recommendation methods will be explored. We will also consider some recommendation methods that have not been used (yet) for a movie recommender.

In addition, this book will propose hybrid recommender systems for movies that use both a content-based (CB) recommendation method and a collaborative filtering (CF) method. By combining these two recommendation methods, we hope to build systems with a higher accuracy of predictions. These methods need to be based on data mining algorithms like neural networks or decision trees. We also hope to improve the prediction quality of recommenders based on these prediction algorithms compared to other systems proposed in the literature. The predictive accuracy of all these recommender systems will be

tested on real life movie data: content information and rating information of movies. The content information will be extracted from a movie and TV site, the Internet Movie Database (IMDb) [1], and the rating information is from Netflix [2], which is an online movie-renting site. Each of these recommenders will predict the number of stars given to the movies by a user, so the prediction can tell to which extent the user will like or dislike the movie.

## **1.2 Motivation**

In our previous work [3] we explored the hybridization method combining the content-based method and the collaborative filtering, both based on the naïve Bayesian classifier. The proposed recommendation methods in that work used two classes: it predicts if a user would like or dislike a movie. In this paper we want to explore hybridization methods that also combine the content-based method and the collaborative filtering, but these methods will be based on neural networks or decision trees. We wanted to explore these combinations, because these combinations has not been researched in the literature before and it will be interesting to see how these hybridization methods will perform compared to other recommendation methods. In addition, in this paper we want to examine the preference of a user more accurate. In other words, we will also look to which extent a user will like or dislike a movie. So instead of only predicting if a user will like or dislike a movie, as we did in our previous work, the prediction will also be divided into five different ratings, from one star till five stars. Here a rating of one star means that the user did not like the movie at all and a rating of five stars means that the user liked the movie very much.

## **1.3 Goal**

In this book we will examine the performance of the proposed hybrid recommender system. The following research question will be answered:

*How does a hybrid recommender for movies based on neural network or decision tree perform, that combines a content-based recommender for movies, which uses text mining, with a collaborative filtering recommender for movies, which uses user ratings?*

To answer the research question, the following sub questions need to be answered first:

1. *How can these two algorithms be used individually for a content-based recommender or collaborative filtering for movies.?*
2. *How can one devise a hybrid recommender based on each of these algorithms, that combines a content-based with a collaborative filtering, both based on one of these algorithms?*

For the first sub question, we will work with content-based and collaborative filtering systems separately, so we will not work with hybrid systems. For both of these recommendation methods we will create a recommender based on neural network and decision tree, so we will have four different recommender systems:

- 1) A content-based recommender system based on neural network (CB-NN).
- 2) A content-based recommender system based on decision tree (CB-DT).
- 3) A collaborative filtering system based on neural network (CF-NN).
- 4) A collaborative filtering system based on decision tree (CF-DT).

The second sub question means that we will work with hybrid recommender systems based on the aforementioned algorithms separately, for both the content-based part and the collaborative filtering part of the system. In addition, we have the following two hybrid recommender systems:



- 5) A hybrid recommender system based on neural network, combining content-based and collaborative filtering (H-NN).
- 6) A hybrid recommender system based on decision tree, combining content-based and collaborative filtering (H-DT)

The performance of the hybrid recommender H-NN will be compared with the recommenders CB-NN and CF-NN and the hybrid recommender H-DT will be compared with the recommender CB-DT and CF-DT. These recommenders will also be compared with the content-based, collaborative filtering and the hybrid recommender systems based on naïve Bayesian classifier.

#### **1.4 Methodology**

In order to answer the research question and the sub questions, we have taken the following steps:

1. *Study literature*
2. *Collect datasets*
3. *Implement algorithms*
4. *Experiment*

##### *1. Study literature*

To answer the research questions, some literature about recommender systems and data mining algorithms have to be studied first. Especially recommender systems for movies will be examined. In the literature, various recommendation methods and algorithms have been discussed. The recommendation methods used in this book are the content-based method and collaborative filtering. The algorithms that are used are neural network and decision tree. Beside these recommendation methods and algorithms, some literature about hybrid recommenders will be studied to find a combination of recommendation methods to improve the prediction of accuracy of the individual methods.