

Transactions on Computer Systems and Networks

Pushpendu Kar
Monideepa Roy
Sujoy Datta

Recommender Systems: Algorithms and their Applications

 Springer

Transactions on Computer Systems and Networks


Series Editor

Amlan Chakrabarti, Director and Professor, A. K. Choudhury School of Information Technology, Kolkata, West Bengal, India

Editorial Board

Jürgen Becker, Institute for Information Processing—ITIV, Karlsruhe Institute of Technology—KIT, Karlsruhe, Germany

Yu-Chen Hu, Department of Computer Science and Information Management, Providence University, Taichung City, Taiwan

Anupam Chattopadhyay , School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore

Gaurav Tribedi, Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Guwahati, India

Sriparna Saha, Department of Computer Science and Engineering, Indian Institute of Technology Patna, Patna, India

Saptarsi Goswami, A. K. Choudhury school of Information Technology, Kolkata, India

Transactions on Computer Systems and Networks is a unique series that aims to capture advances in evolution of computer hardware and software systems and progress in computer networks. Computing Systems in present world span from miniature IoT nodes and embedded computing systems to large-scale cloud infrastructures, which necessitates developing systems architecture, storage infrastructure and process management to work at various scales. Present day networking technologies provide pervasive global coverage on a scale and enable multitude of transformative technologies. The new landscape of computing comprises of self-aware autonomous systems, which are built upon a software-hardware collaborative framework. These systems are designed to execute critical and non-critical tasks involving a variety of processing resources like multi-core CPUs, reconfigurable hardware, GPUs and TPUs which are managed through virtualisation, real-time process management and fault-tolerance. While AI, Machine Learning and Deep Learning tasks are predominantly increasing in the application space the computing system research aim towards efficient means of data processing, memory management, real-time task scheduling, scalable, secured and energy aware computing. The paradigm of computer networks also extends its support to this evolving application scenario through various advanced protocols, architectures and services. This series aims to present leading works on advances in theory, design, behaviour and applications in computing systems and networks. The Series accepts research monographs, introductory and advanced textbooks, professional books, reference works, and select conference proceedings.

Pushpendu Kar · Monideepa Roy · Sujoy Datta

Recommender Systems: Algorithms and their Applications

Pushpendu Kar
School of Computer Science
University of Nottingham Ningbo China
Ningbo, China

Monideepa Roy
School of Computer Engineering
KIIT Deemed University
Bhubaneswar, Odisha, India

Sujoy Datta
School of Computer Engineering
KIIT Deemed University
Bhubaneswar, Odisha, India

ISSN 2730-7484 ISSN 2730-7492 (electronic)
Transactions on Computer Systems and Networks
ISBN 978-981-97-0537-5 ISBN 978-981-97-0538-2 (eBook)
<https://doi.org/10.1007/978-981-97-0538-2>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

If disposing of this product, please recycle the paper.

Preface

Recommendation systems were introduced in the 90's but have gradually become an indispensable tool with the advent of numerous e-commerce companies. Recent years have seen a huge jump in the number of such web services, and they rely heavily on recommendation systems to gain an advantage over their competitors. Recommendation systems gather information about the likes and dislikes of a user and use various types of complex algorithms to predict what a user may be interested in and send personalized recommendations to users. Brands like Netflix, Amazon, Facebook, Spotify, and YouTube collect information about users and try to predict user preferences. If a person buys a certain product, then suggestions for similar products are sent to the user. If a user likes a particular type of music or movie, then it will try to predict and recommend similar types of music or movies to the user. It is a very vast and interesting area of research but at present, in this book, we have taken some of the most important topics which form the basis of recommender systems, along with some case studies and applications and suggestions for future research directions.

This book will be useful to users who are new to the topic and wish to learn it. It will also be useful to advanced users who know the theory but want to implement or design a system from scratch and can learn from the different types of algorithms.

This book consists of 12 chapters.

Chapter 1 is a general introduction of what is the importance of recommender systems and an overview of the scope of the book and its audience and the motivation behind writing this book.

Chapter 2 is a general overview of all possible types of algorithms for recommendation systems.

Chapter 3 discusses two of the most widely used types of recommender algorithms, content-based systems and collaborative filtering methods, and their features and suitability for implementation.

Chapter 4 discusses the decomposition of the matrix in clustering.

Chapter 5 discusses how to learn to rank users based on various factors and how to detect profiles of false users, along with the Shilling attack example.

Chapter 6 deals with knowledge-based, ensemble-based, and hybrid recommender systems.

Chapter 7 discusses how to deal with the big data associated with recommender systems.

Chapter 8 discusses the existing trust-centric and attack-resistance techniques for recommender systems and proposes different ways to improve the performance of recommendation systems based on both attack and trust.

Chapter 9 shows the steps in building a recommendation engine.

Chapter 10 discusses different types of healthcare recommendation systems, challenges, and the scope of improvements.

Chapter 11 discusses the application of recommender systems to military surveillance.

Chapter 12 discusses the use of recommender systems in different real application domains, existing challenges as well as the scopes and ideas of their improvements.

Ningbo, China
Bhubaneswar, India
Bhubaneswar, India

Pushpendu Kar
Monideepa Roy
Sujoy Datta

Acknowledgments

I would like to thank my parents, **Mihir Kumar Kar** and **Pratima Kar**, my wife, **Sangita**, my daughter, **Ritosmita**, and my son, **Ritanshu**, for their continuous support, guidance, and encouragement. I would like to express my sincere gratitude to **Tianyi Ma** and **Chenyu Yang** for helping in writing Chap. 8, **Zhihang Zhu** for helping in writing Chap. 10, and **Xinyi Wang** for helping in writing Chap. 12.

—Pushpendu Kar

I would like to express my sincere thanks to my Mom (**late Hasi Roy**) and Dad (**late Sunil K. Roy**) for their blessings even when they are no more physically here to guide me, but I'm sure they are watching with satisfaction from above. I would like to thank my sister **Madhumita Roy** for her constant support and inspiration throughout this assignment. I am also thankful to her for kindly designing the cover for the book. I would like to thank my B.Tech. students **Aishi Paul** and **Divyansi Mishra** for their help in drawing the diagrams for the book. Thanks to each and every one of you for your timely support and help.

—Monideepa Roy

I would like to express my sincere thanks to my father (**Sushil Dutta**) and my mother (**late Mintu Dutta**) for their constant inspiration and blessings. I would also like to thank my sisters **Sutapa Rakshit** and **Sunanda Rakshit** for their continuous encouragement and support to me during this assignment.

—Sujoy Datta

Contents

1	Introduction to Recommendation Systems	1
1.1	Introduction	1
1.2	What Are Recommendation Systems?	2
1.3	Who Can Benefit from Them?	2
1.4	How Do They Help?	4
1.5	How Do They Work?	5
1.6	The Evolution of Recommender Systems	5
1.7	Some Major Brands Who Are Very Successfully Based on Recommendation Systems	7
1.8	Scope of the Book	9
1.9	Overview of the Chapters	9
1.10	Summary	9
	References	10
2	Overview of Recommendation Systems	11
2.1	Introduction	11
2.2	Goals of Recommender Systems	12
2.3	The Spectrum of Applications	12
2.4	Classification of Recommendation Systems	13
2.5	Domain-Specific Challenges in Recommender Systems	15
2.6	Summary	16
	References	16
3	Collaborative Filtering and Content-Based Systems	19
3.1	Introduction	19
3.2	Collaborative Filtering Model	20
3.3	Content-Based Recommender Systems	22
3.4	Hybrid Recommender Systems	23
3.5	Similarity Measures Used by a Recommender System	25
3.5.1	Distance-Based Similarity Measure	25
3.5.2	Correlation-Based Similarity	26

- 3.6 Evaluation Metrics of Recommender Systems (Ge et al. 2010) 27
 - 3.6.1 Mean Absolute Error (MAE) 27
 - 3.6.2 Root Mean Square Error (RMSE) 27
 - 3.6.3 Precision 27
 - 3.6.4 Recall 28
 - 3.6.5 F₁ Score and Accuracy 28
- 3.7 Summary 28
- References 29
- 4 Matrix Decomposition for Clustering and Collaborative Filtering** 31
 - 4.1 Introduction 31
 - 4.2 Why Matrix Decomposition is Useful? 32
 - 4.3 The Matrix Decomposition Technique 33
 - 4.4 The Netflix Example 35
 - 4.5 Summary 37
 - References 38
- 5 Learning How to Rank and Collecting User Behavior** 39
 - 5.1 Introduction 39
 - 5.2 The Facebook and FourSquare (Factual) Ranking Methods 40
 - 5.3 Feature Selection in Recommender Systems 41
 - 5.4 The Ranking Module in a Recommender System 46
 - 5.5 Introduction to Ranking Algorithms 47
 - 5.6 Collecting User Likes and Dislikes 48
 - 5.7 Detection of Fake/Malicious Profiles 49
 - 5.8 Shilling/Profile Injection Attacks 50
 - 5.9 Detection Algorithms 52
 - 5.10 Summary 53
 - References 54
- 6 Knowledge-Based, Ensemble-Based, and Hybrid Recommender Systems** 55
 - 6.1 Introduction 55
 - 6.2 The Cold Start Problem 56
 - 6.3 Knowledge-Based Recommender Systems 57
 - 6.3.1 Constraint-Based Recommender Systems 59
 - 6.3.2 Case-Based Recommender Systems 61
 - 6.4 Ensemble-Based and Hybrid Recommender Systems 63
 - 6.5 Ensemble Methods from the Classification Perspective 65
 - 6.6 Summary 66
 - References 67

- 7 Big Data Behind Recommender Systems** 69
 - 7.1 Introduction 69
 - 7.2 What is Big Data? 70
 - 7.3 How Big Data is Used in Recommender Systems 71
 - 7.4 Types of Data Used in Recommender Systems 72
 - 7.5 Challenges 73
 - 7.6 An Example of the Role of Big Data in Twitter 75
 - 7.7 Singular Value Decomposition-Based Recommender Systems 76
 - 7.7.1 Singular Value Decomposition (SVD) 76
 - 7.7.2 Recommender Systems Using SVD 77
 - 7.8 Summary 79
 - References 79

- 8 Trust-Centric and Attack-Resistant Recommender System** 81
 - 8.1 Introduction 81
 - 8.2 Literature Review 82
 - 8.2.1 Concept of Trust in Recommender Systems 82
 - 8.2.2 Social Scoring 85
 - 8.2.3 Evolution of Trust-Based Recommender Systems 86
 - 8.2.4 Concept of Attack in Recommender System 87
 - 8.3 Challenges in Previous Research 90
 - 8.3.1 Challenges in Researches About Trust 90
 - 8.3.2 Challenges in Researches About Attack 90
 - 8.4 Possible Improvements for Future Research 92
 - 8.4.1 Improvements in Score Propagation Model 92
 - 8.4.2 User’s Activeness 94
 - 8.4.3 Administrative Measures 97
 - 8.5 Summary 99
 - References 100

- 9 Steps in Building a Recommendation Engine** 101
 - 9.1 Introduction 101
 - 9.2 Design/Evaluation Parameters of a Recommender System 102
 - 9.3 Overview of the Ways to Design a Recommendation System ... 105
 - 9.4 Steps in Building a Successful Recommendation Engine 108
 - 9.4.1 Understanding the Business 108
 - 9.4.2 Getting the Data 108
 - 9.4.3 Explore, Clean and Augment the Data 108
 - 9.4.4 Predict the Ranking 109
 - 9.4.5 Visualizing the Data 109
 - 9.4.6 Iterate and Deploy the Models 109
 - 9.5 Architecture 109
 - 9.6 Summary 110
 - References 111

- 10 Recommender System for Health Care** 113
 - 10.1 Introduction 113
 - 10.2 Analysis of Healthcare Recommender System 114
 - 10.2.1 HRS for Diagnosis 115
 - 10.2.2 HRS for Medicine 120
 - 10.2.3 HRS for Lifestyle 126
 - 10.3 Real Example 129
 - 10.4 Summary 129
 - References 129
- 11 A Surveillance Framework of Suspicious Browsing Activities on the Internet Using Recommender Systems: A Case Study** 131
 - 11.1 Introduction 131
 - 11.2 Related Work 132
 - 11.3 Web User Tracking of Browsing Patterns and Populating Recommender Systems 133
 - 11.4 Sparse Matrices 135
 - 11.5 Our Proposed Framework 136
 - 11.6 The Proposed Algorithm for the Threat Analysis and Alert 138
 - 11.7 Real Example 140
 - 11.8 Summary 140
 - References 140
- 12 Some Novel Applications of Recommender System and Road Ahead** 143
 - 12.1 Introduction 143
 - 12.2 Applications of Recommender System 144
 - 12.2.1 Security 144
 - 12.2.2 Tourism 146
 - 12.2.3 E-commerce 150
 - 12.2.4 E-learning 154
 - 12.2.5 Social Network 158
 - 12.3 Summary 160
 - References 161
- Index** 163

About the Authors



Dr. Pushpendu Kar is currently working as an Assistant Professor in the School of Computer Science, University of Nottingham (China campus). Before this, he was a Postdoctoral Research Fellow at the Norwegian University of Science and Technology, the National University of Singapore, and Nanyang Technological University. He also worked in different engineering colleges as a lecturer and in the IT industry as a software professional. He has more than 12 years of teaching and research experience as well as one and a half years of industrial experience at IBM. He has completed all his Ph.D., Master of Engineering, and Bachelor of Technology in Computer Science and Engineering. He also completed Postgraduate Certificate in Higher Education (PGCHE) in 2023. He was awarded the prestigious Erasmus Mundus Postdoctoral Fellowship from the European Commission, ERCIM Alain Bensoussan Fellowship from the European Union, and SERB OPD Fellowship from the Department of Science and Technology, Government of India. He has received the 2020 IEEE Systems Journal Best Paper Award. He has received four research grants for conducting research-based projects, three of them as a Principal Investigator (PI). He also received many travel grants to attend conferences and doctoral colloquiums. He is the author of more than 60 scholarly research papers, which have been published in reputed journals, conferences, book chapters, and IT magazines. He has also published three books. He is an inventor of five patents. He has chaired several conference committees, worked as a team member to organize short-term

courses, and delivered a few invited talks as well as Keynote Lectures at international conferences and institutions. He is a Senior Member of IEEE, a Senior Fellow of the Higher Education Academy (SFHEA), UK, and Fellow of the Institution of Electronics and Telecommunication Engineers (FIETE), India. Ningbo Municipal Government, China has recognized him as a High-Level Talent. His research areas include Wireless Sensor Networks, Internet of Things, Content-Centric Networking, Machine Learning, and Blockchain.



Dr. Monideepa Roy did her Bachelors and Masters in Mathematics from IIT Kharagpur, and her Ph.D. in CSE from Jadavpur University. Currently she is working as an Associate Professor at KIIT Deemed University, Bhubaneswar since the last 11 years. Her areas of interest include Remote Healthcare, Mobile Computing, Cognitive WSNs, Remote Sensing, Recommender Systems, Sparse Approximations, and Artificial Neural Networks. At present she has seven research scholars working with her in the above areas and two more have successfully defended their theses under her guidance. She has several publications in reputed conferences and journals. She has been the Organizing Chair of the first two editions of the International Conference on Computational Intelligence and Networks CINE 2015 and 2016, ICMC 2019 and has organised several workshops and seminars. She also has several book chapter publications in Springer as well as two edited books under Taylor and Francis.



Sujoy Datta has done his M.Tech. from IIT Kharagpur. Currently he is working as an Assistant Professor in the School of Computer Engineering, KIIT Deemed University since the last eleven years. His areas of research include Wireless networks, Computer Security, Elliptic curve cryptography and neural networks, Remote Healthcare and Recommender Systems. He has several publications in various conferences and journals. He has co-organised several workshops and international conferences as well as several workshops and seminars. He also has several book chapter publications in Springer as well as two edited books by Taylor and Francis.

Chapter 1

Introduction to Recommendation Systems



Abstract With the rapid growth of e-commerce, the web has become a very popular source of doing business by various companies. Customers also find it a very attractive proposition as it saves the time to go outside and shop for what a user needs, as well as the fact that users have access to a huge array of choices to buy from. Since it is a very tough and competitive market, and companies have realized that people usually tend to buy similar types of products or watch similar types of movies, they have now resorted to modern technology to make it easier for customers to make their choices. This led to the advent of various recommendation algorithms with which the companies are now able to predict the choices and personal preferences of their customers and accordingly push appropriate suggestions or recommendations for products that a person is likely to purchase. With the huge success of recommendation systems, they are now widely being adopted by more and more brands and for more varieties of applications. This chapter gives an overview of the reasons why recommendation systems have become so popular.

Keywords Recommender algorithms · Product recommendations · Machine learning · Customer choices · Prediction

1.1 Introduction

Since consumers today are faced with huge numbers of choices in terms of new products or new movies to watch, and less time on their hands, so it's difficult to make the choices of selection of the most relevant options on their own. So, whenever a person buys a new product or wants to watch a new movie, he/she prefers to find out ratings or recommendations from past users to make their choices faster and easier (Abowd et al. 1999). However, even that is time-consuming with the huge volumes of data. So, this has led to the emergence of recommendation systems, which use algorithms to predict and find the best matches for a person based on various parameters. They also form the basis of many machine learning algorithms. In this book, we take a look at the different types of algorithms that are used for generating

accurate predictions for consumers and some applications of recommender systems. The rest of the chapter is organized as follows: Sect. 1.2 defines what recommendation systems are, Sect. 1.3 discusses who can benefit from recommendation systems, Sect. 1.4 discusses how that can help, Sect. 1.5 describes how they work, Sect. 1.6 describes the evolution of recommender systems, Sect. 1.7 takes the examples of some famous brands who have used recommendation systems very successfully, Sect. 1.8 describes the scope of the book, Sect. 1.9 gives an overview of the chapters, and Sect. 1.10 is the summary.

1.2 What Are Recommendation Systems?

As defined in Wikipedia, “A recommender system, or a recommendation system (sometimes replacing ‘system’ with a synonym such as a platform or an engine), is a subclass of information filtering system that seeks to predict the “rating” or “preference” a user would give to an item.”

So, the basic aim of a recommender system is to provide users with the most relevant suggestions for things to buy, places to visit, or movies to watch, based on what the user had chosen earlier or what people with similar profiles have chosen. Recommender systems (Adamopoulos et al. 2014) predict the choices of the users and then suggest the most relevant options. In this present age of competition, there is a huge number of choices available to consumers, especially in the e-commerce domain. So to gain an edge over other competitors, a retailer or business needs to be able to correctly predict user preferences and send appropriate suggestions for their products to the consumers. This makes recommendation systems the most powerful machine learning technique which is widely used by online retailers to gain an edge over the others and increase their profits. So how do they actually work and how do they predict the preferences of their users? The data that is required for the recommendation systems to make such predictions is obtained from various sources. Data is collected explicitly from user ratings that are collected after a person has watched a movie or listened to a song or purchased an item, implicitly through search engine queries and purchase histories through cookies, or from past knowledge about the user or the item. Many sites like Netflix, Facebook, Amazon, Spotify, and YouTube use such types of data and implement their algorithms to provide the most relevant suggestions to the users.

1.3 Who Can Benefit from Them?

Although any business can benefit from implementing recommendation systems, the two main factors which determine the extent to which a business can benefit from recommendation systems are:

Breadth of data—If the business has only a few customers, and they behave in different ways, then using an automated recommendation system will not be of much use to them. It will be much easier to let the employees use their own logic to predict the preferences of the individual customers.

Depth of data—When the business has only one single data point for each of their customers, recommendation systems will not have sufficient training data to base their predictions on.

So organizations who can benefit from automated recommendation systems can vary from e-commerce, retail and media to banking, telecom and other utilities. There are of course many more areas which can benefit from implementing recommendation systems, but here we describe some of the most popular ones. A more detailed discussion of specific brands will be done in the later chapters.

E-commerce is one of the first areas where recommendation systems were used. Because these companies have access to online data of millions of customers, they can easily use that data to generate accurate recommendations.

Retail is another area which can benefit to a great extent from recommendation systems. Since retailers have direct access to huge volumes of shopping data therefore they have a very good idea of the customers intent and can make accurate predictions.

Media industry is also one of the first few companies which were the first to implement recommendation systems. Almost all news channels use recommendation engines.

Banking is also a very important application where the financial situations and past preferences of millions of customer data make a very comprehensive data bank.

The telecom industry also has a similar dynamics as that of the banking industry where the service providers have access to a wide variety of customer data, call and usage preferences, and past data of a huge volume of customers. The telecom industry has the additional advantage that it has a limited number of products for which they need information, to make their predictions on (Fig. 1.1).



Fig. 1.1 Increasing importance of personalization in the post pandemic market. Source - McKinsey