

Recommender Systems Handbook

Francesco Ricci · Lior Rokach · Bracha Shapira ·
Paul B. Kantor
Editors

Recommender Systems Handbook

 Springer

Editors

Francesco Ricci
Free University of Bozen-Bolzano
Faculty of Computer Science
Piazza Domenicani 3
39100 Bolzano
Italy
fricci@unibz.it

Lior Rokach
Ben-Gurion University of the
Negev
Dept. Information Systems
Engineering
84105 Beer-Sheva
Israel
liorrk@bgu.ac.il

Bracha Shapira
Ben-Gurion University of the
Negev
Dept. Information Systems
Engineering
Beer-Sheva
Israel
bshapira@bgu.ac.il

Paul B. Kantor
Rutgers University
School of Communication,
Information & Library Studies
Huntington Street 4
08901-1071 New Brunswick
New Jersey
SCILS Bldg.
USA
kantor@scils.rutgers.edu

ISBN 978-0-387-85819-7 e-ISBN 978-0-387-85820-3
DOI 10.1007/978-0-387-85820-3
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010937590

© Springer Science+Business Media, LLC 2011

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

*Dedicated to our families in appreciation for
their patience and support during the
preparation of this handbook.*

F.R.

L.R.

B.S.

P.K.

Preface

Recommender Systems are software tools and techniques providing suggestions for items to be of use to a user. The suggestions provided are aimed at supporting their users in various decision-making processes, such as what items to buy, what music to listen, or what news to read. Recommender systems have proven to be valuable means for online users to cope with the information overload and have become one of the most powerful and popular tools in electronic commerce. Correspondingly, various techniques for recommendation generation have been proposed and during the last decade, many of them have also been successfully deployed in commercial environments.

Development of recommender systems is a multi-disciplinary effort which involves experts from various fields such as Artificial intelligence, Human Computer Interaction, Information Technology, Data Mining, Statistics, Adaptive User Interfaces, Decision Support Systems, Marketing, or Consumer Behavior. *Recommender Systems Handbook: A Complete Guide for Research Scientists and Practitioners* aims to impose a degree of order upon this diversity by presenting a coherent and unified repository of recommender systems' major concepts, theories, methodologies, trends, challenges and applications. This is the first comprehensive book which is dedicated entirely to the field of recommender systems and covers several aspects of the major techniques. Its informative, factual pages will provide researchers, students and practitioners in industry with a comprehensive, yet concise and convenient reference source to recommender systems. The book describes in detail the classical methods, as well as extensions and novel approaches that were recently introduced. The book consists of five parts: techniques, applications and evaluation of recommender systems, interacting with recommender systems, recommender systems and communities, and advanced algorithms. The first part presents the most popular and fundamental techniques used nowadays for building recommender systems, such as collaborative filtering, content-based filtering, data mining methods and context-aware methods. The second part starts by surveying techniques and approaches that have been used to evaluate the quality of the recommendations. Then deals with the practical aspects of designing recommender systems, it describes design and implementation consideration, setting guidelines for the selection of the

more suitable algorithms. The section continues considering aspects that may affect the design and finally, it discusses methods, challenges and measures to be applied for the evaluation of the developed systems. The third part includes papers dealing with a number of issues related to the presentation, browsing, explanation and visualization of the recommendations, and techniques that make the recommendation process more structured and conversational.

The fourth part is fully dedicated to a rather new topic, which is however rooted in the core idea of a collaborative recommender, i.e., exploiting user generated content of various types to build new types and more credible recommendations.

Finally the last section collects a few papers on some advanced topics, such as the exploitation of active learning principles to guide the acquisition of new knowledge, techniques suitable for making a recommender system robust against attacks of malicious users, and recommender systems that aggregate multiple types of user feedbacks and preferences to build more reliable recommendations.

We would like to thank all authors for their valuable contributions. We would like to express gratitude for all reviewers that generously gave comments on drafts or counsel otherwise. We would like to express our special thanks to Susan Lagerstrom-Fife and staff members of Springer for their kind cooperation throughout the production of this book. Finally, we wish this handbook will contribute to the growth of this subject, we wish to the novices a fruitful learning path, and to those more experts a compelling application of the ideas discussed in this handbook and a fruitful development of this challenging research area.

Francesco Ricci
Lior Rokach
Bracha Shapira
Paul B. Kantor

May 2010

Contents

1	Introduction to Recommender Systems Handbook	1
	Francesco Ricci, Lior Rokach and Bracha Shapira	
1.1	Introduction	1
1.2	Recommender Systems Function	4
1.3	Data and Knowledge Sources	7
1.4	Recommendation Techniques	10
1.5	Application and Evaluation	14
1.6	Recommender Systems and Human Computer Interaction	17
1.6.1	Trust, Explanations and Persuasiveness	18
1.6.2	Conversational Systems	19
1.6.3	Visualization	21
1.7	Recommender Systems as a Multi-Disciplinary Field	21
1.8	Emerging Topics and Challenges	23
1.8.1	Emerging Topics Discussed in the Handbook	23
1.8.2	Challenges	26
	References	29

Part I Basic Techniques

2	Data Mining Methods for Recommender Systems	39
	Xavier Amatriain, Alejandro Jaimes, Nuria Oliver, and Josep M. Pujol	
2.1	Introduction	39
2.2	Data Preprocessing	40
2.2.1	Similarity Measures	41
2.2.2	Sampling	42
2.2.3	Reducing Dimensionality	44
2.2.4	Denoising	47
2.3	Classification	48
2.3.1	Nearest Neighbors	48
2.3.2	Decision Trees	50
2.3.3	Ruled-based Classifiers	51

2.3.4	Bayesian Classifiers	52
2.3.5	Artificial Neural Networks	54
2.3.6	Support Vector Machines	56
2.3.7	Ensembles of Classifiers	58
2.3.8	Evaluating Classifiers	59
2.4	Cluster Analysis	61
2.4.1	k -Means	62
2.4.2	Alternatives to k -means	63
2.5	Association Rule Mining	64
2.6	Conclusions	66
	References	67
3	Content-based Recommender Systems: State of the Art and Trends	73
	Pasquale Lops, Marco de Gemmis and Giovanni Semeraro	
3.1	Introduction	74
3.2	Basics of Content-based Recommender Systems	75
3.2.1	A High Level Architecture of Content-based Systems	75
3.2.2	Advantages and Drawbacks of Content-based Filtering	78
3.3	State of the Art of Content-based Recommender Systems	79
3.3.1	Item Representation	80
3.3.2	Methods for Learning User Profiles	90
3.4	Trends and Future Research	94
3.4.1	The Role of User Generated Content in the Recommendation Process	94
3.4.2	Beyond Over-specialization: Serendipity	96
3.5	Conclusions	99
	References	100
4	A Comprehensive Survey of Neighborhood-based Recommendation Methods	107
	Christian Desrosiers and George Karypis	
4.1	Introduction	107
4.1.1	Formal Definition of the Problem	108
4.1.2	Overview of Recommendation Approaches	110
4.1.3	Advantages of Neighborhood Approaches	112
4.1.4	Objectives and Outline	113
4.2	Neighborhood-based Recommendation	114
4.2.1	User-based Rating Prediction	115
4.2.2	User-based Classification	116
4.2.3	Regression VS Classification	117
4.2.4	Item-based Recommendation	117
4.2.5	User-based VS Item-based Recommendation	118
4.3	Components of Neighborhood Methods	120
4.3.1	Rating Normalization	121
4.3.2	Similarity Weight Computation	124
4.3.3	Neighborhood Selection	129

4.4	Advanced Techniques	131
4.4.1	Dimensionality Reduction Methods	132
4.4.2	Graph-based Methods	135
4.5	Conclusion	139
	References	140
5	Advances in Collaborative Filtering	145
	Yehuda Koren and Robert Bell	
5.1	Introduction	145
5.2	Preliminaries	147
5.2.1	Baseline predictors	148
5.2.2	The Netflix data	149
5.2.3	Implicit feedback	150
5.3	Matrix factorization models	151
5.3.1	SVD	151
5.3.2	SVD++	153
5.3.3	Time-aware factor model	154
5.3.4	Comparison	159
5.3.5	Summary	160
5.4	Neighborhood models	161
5.4.1	Similarity measures	162
5.4.2	Similarity-based interpolation	163
5.4.3	Jointly derived interpolation weights	165
5.4.4	Summary	168
5.5	Enriching neighborhood models	168
5.5.1	A global neighborhood model	169
5.5.2	A factorized neighborhood model	173
5.5.3	Temporal dynamics at neighborhood models	180
5.5.4	Summary	182
5.6	Between neighborhood and factorization	182
	References	184
6	Developing Constraint-based Recommenders	187
	Alexander Felfernig, Gerhard Friedrich, Dietmar Jannach and Markus Zanker	
6.1	Introduction	187
6.2	Development of Recommender Knowledge Bases	191
6.3	User Guidance in Recommendation Processes	194
6.4	Calculating Recommendations	203
6.5	Experiences from Projects and Case Studies	205
6.6	Future Research Issues	207
6.7	Summary	212
	References	212

7	Context-Aware Recommender Systems	217
	Gediminas Adomavicius and Alexander Tuzhilin	
7.1	Introduction and Motivation	218
7.2	Context in Recommender Systems	219
7.2.1	What is Context?	219
7.2.2	Modeling Contextual Information in Recommender Systems	223
7.2.3	Obtaining Contextual Information	228
7.3	Paradigms for Incorporating Context in Recommender Systems	230
7.3.1	Contextual Pre-Filtering	233
7.3.2	Contextual Post-Filtering	237
7.3.3	Contextual Modeling	238
7.4	Combining Multiple Approaches	243
7.4.1	Case Study of Combining Multiple Pre-Filters: Algorithms	244
7.4.2	Case Study of Combining Multiple Pre-Filters: Experimental Results	245
7.5	Additional Issues in Context-Aware Recommender Systems	247
7.6	Conclusions	249
	References	250

Part II Applications and Evaluation of RSs

8	Evaluating Recommendation Systems	257
	Guy Shani and Asela Gunawardana	
8.1	Introduction	258
8.2	Experimental Settings	260
8.2.1	Offline Experiments	261
8.2.2	User Studies	263
8.2.3	Online Evaluation	266
8.2.4	Drawing Reliable Conclusions	267
8.3	Recommendation System Properties	271
8.3.1	User Preference	272
8.3.2	Prediction Accuracy	273
8.3.3	Coverage	281
8.3.4	Confidence	283
8.3.5	Trust	285
8.3.6	Novelty	285
8.3.7	Serendipity	286
8.3.8	Diversity	288
8.3.9	Utility	289
8.3.10	Risk	290
8.3.11	Robustness	290
8.3.12	Privacy	291
8.3.13	Adaptivity	292

8.3.14	Scalability	293
8.4	Conclusion	293
	References	294
9	A Recommender System for an IPTV Service Provider: a Real Large-Scale Production Environment.	299
	Riccardo Bambini, Paolo Cremonesi and Roberto Turrin	
9.1	Introduction	299
9.2	IPTV Architecture	301
9.2.1	IPTV Search Problems	302
9.3	Recommender System Architecture	303
9.3.1	Data Collection	304
9.3.2	Batch and Real-Time Stages	306
9.4	Recommender Algorithms	308
9.4.1	Overview of Recommender Algorithms	308
9.4.2	LSA Content-Based Algorithm	311
9.4.3	Item-based Collaborative Algorithm	314
9.4.4	Dimensionality-Reduction-Based Collaborative Algorithm	316
9.5	Recommender Services	318
9.6	System Evaluation	319
9.6.1	Off-Line Analysis	321
9.6.2	On-line Analysis	325
9.7	Conclusions	329
	References	329
10	How to Get the Recommender Out of the Lab?	333
	J��r��me Picault, Myriam Rib��re, David Bonnefoy and Kevin Mercer	
10.1	Introduction	334
10.2	Designing Real-World Recommender Systems	334
10.3	Understanding the Recommender Environment	335
10.3.1	Application Model	335
10.3.2	User Model	340
10.3.3	Data Model	344
10.3.4	A Method for Using Environment Models	349
10.4	Understanding the Recommender Validation Steps in an Iterative Design Process	350
10.4.1	Validation of the Algorithms	350
10.4.2	Validation of the Recommendations	351
10.5	Use Case: a Semantic News Recommendation System	355
10.5.1	Context: MESH Project	356
10.5.2	Environmental Models in MESH	357
10.5.3	In Practice: Iterative Instantiations of Models	361
10.6	Conclusion	362
	References	362

11	Matching Recommendation Technologies and Domains	367
	Robin Burke and Maryam Ramezani	
11.1	Introduction	367
11.2	Related Work	368
11.3	Knowledge Sources	368
11.3.1	Recommendation types	370
11.4	Domain	372
11.4.1	Heterogeneity	372
11.4.2	Risk	373
11.4.3	Churn	373
11.4.4	Interaction Style	374
11.4.5	Preference stability	374
11.4.6	Scrutability	375
11.5	Knowledge Sources	375
11.5.1	Social Knowledge	375
11.5.2	Individual	376
11.5.3	Content	377
11.6	Mapping Domains to Technologies	378
11.6.1	Algorithms	380
11.6.2	Sample Recommendation Domains	381
11.7	Conclusion	382
	References	382
12	Recommender Systems in Technology Enhanced Learning	387
	Nikos Manouselis, Hendrik Drachsler, Riina Vuorikari, Hans Hummel and Rob Koper	
12.1	Introduction	388
12.2	Background	389
12.3	Related Work	392
12.4	Survey of TEL Recommender Systems	399
12.5	Evaluation of TEL Recommenders	404
12.6	Conclusions and further work	408
	References	409

Part III Interacting with Recommender Systems

13	On the Evolution of Critiquing Recommenders	419
	Lorraine McGinty and James Reilly	
13.1	Introduction	419
13.2	The Early Days: Critiquing Systems/Recognised Benefits	420
13.3	Representation & Retrieval Challenges for Critiquing Systems	422
13.3.1	Approaches to Critique Representation	422
13.3.2	Retrieval Challenges in Critique-Based Recommenders	430
13.4	Interfacing Considerations Across Critiquing Platforms	438
13.4.1	Scaling to Alternate Critiquing Platforms	438
13.4.2	Direct Manipulation Interfaces vs Restricted User Control	440

13.4.3	Supporting Explanation, Confidence & Trust	441
13.4.4	Visualisation, Adaptivity, and Partitioned Dynamicity . . .	443
13.4.5	Respecting Multi-cultural Usability Differences	445
13.5	Evaluating Critiquing: Resources, Methodologies and Criteria . . .	445
13.5.1	Resources & Methodologies	446
13.5.2	Evaluation Criteria	446
13.6	Conclusion / Open Challenges & Opportunities	448
	References	449
14	Creating More Credible and Persuasive Recommender Systems: The Influence of Source Characteristics on Recommender System Evaluations	455
	Kyung-Hyan Yoo and Ulrike Gretzel	
14.1	Introduction	455
14.2	Recommender Systems as Social Actors	456
14.3	Source Credibility	457
14.3.1	Trustworthiness	458
14.3.2	Expertise	458
14.3.3	Influences on Source Credibility	458
14.4	Source Characteristics Studied in Human-Human Interactions . . .	459
14.4.1	Similarity	459
14.4.2	Likeability	460
14.4.3	Symbols of Authority	460
14.4.4	Styles of Speech	461
14.4.5	Physical Attractiveness	461
14.4.6	Humor	461
14.5	Source Characteristics in Human-Computer Interactions	462
14.6	Source Characteristics in Human-Recommender System Interactions	463
14.6.1	Recommender system type	463
14.6.2	Input characteristics	464
14.6.3	Process characteristics	465
14.6.4	Output characteristics	465
14.6.5	Characteristics of embodied agents	467
14.7	Discussion	468
14.8	Implications	468
14.9	Directions for future research	470
	References	471
15	Designing and Evaluating Explanations for Recommender Systems	479
	Nava Tintarev and Judith Masthoff	
15.1	Introduction	479
15.2	Guidelines	481
15.3	Explanations in Expert Systems	481
15.4	Defining Goals	482
15.4.1	Explain How the System Works: Transparency	483

- 15.4.2 Allow Users to Tell the System it is Wrong: Scrutability 485
- 15.4.3 Increase Users' Confidence in the System: Trust 485
- 15.4.4 Convince Users to Try or Buy: Persuasiveness 487
- 15.4.5 Help Users Make Good Decisions: Effectiveness 488
- 15.4.6 Help Users Make Decisions Faster: Efficiency 490
- 15.4.7 Make the use of the system enjoyable: Satisfaction 491
- 15.5 Evaluating the Impact of Explanations on the Recommender System 492
 - 15.5.1 Accuracy Metrics 493
 - 15.5.2 Learning Rate 493
 - 15.5.3 Coverage 494
 - 15.5.4 Acceptance 494
- 15.6 Designing the Presentation and Interaction with Recommendations 495
 - 15.6.1 Presenting Recommendations 495
 - 15.6.2 Interacting with the Recommender System 496
- 15.7 Explanation Styles 497
 - 15.7.1 Collaborative-Based Style Explanations 500
 - 15.7.2 Content-Based Style Explanation 501
 - 15.7.3 Case-Based Reasoning (CBR) Style Explanations 503
 - 15.7.4 Knowledge and Utility-Based Style Explanations 504
 - 15.7.5 Demographic Style Explanations 505
- 15.8 Summary and future directions 505
- References 507

- 16 Usability Guidelines for Product Recommenders Based on Example Critiquing Research 511**
 - Pearl Pu, Boi Faltings, Li Chen, Jiyong Zhang and Paolo Viappiani
 - 16.1 Introduction 512
 - 16.2 Preliminaries 513
 - 16.2.1 Interaction Model 513
 - 16.2.2 Utility-Based Recommenders 515
 - 16.2.3 The Accuracy, Confidence, Effort Framework 517
 - 16.2.4 Organization of this Chapter 518
 - 16.3 Related Work 518
 - 16.3.1 Types of Recommenders 518
 - 16.3.2 Rating-based Systems 519
 - 16.3.3 Case-based Systems 519
 - 16.3.4 Utility-based Systems 519
 - 16.3.5 Critiquing-based Systems 520
 - 16.3.6 Other Design Guidelines 520
 - 16.4 Initial Preference Elicitation 521
 - 16.5 Stimulating Preference Expression with Examples 525
 - 16.5.1 How Many Examples to Show 527
 - 16.5.2 What Examples to Show 527
 - 16.6 Preference Revision 530

16.6.1	Preference Conflicts and Partial Satisfaction	531
16.6.2	Tradeoff Assistance	532
16.7	Display Strategies	534
16.7.1	Recommending One Item at a Time	534
16.7.2	Recommending K best Items	535
16.7.3	Explanation Interfaces	536
16.8	A Model for Rationalizing the Guidelines	537
16.9	Conclusion	541
	References	541
17	Map Based Visualization of Product Catalogs	547
	Martijn Kagie, Michiel van Wezel and Patrick J.F. Groenen	
17.1	Introduction	547
17.2	Methods for Map Based Visualization	549
17.2.1	Self-Organizing Maps	550
17.2.2	Treemaps	551
17.2.3	Multidimensional Scaling	553
17.2.4	Nonlinear Principal Components Analysis	553
17.3	Product Catalog Maps	554
17.3.1	Multidimensional Scaling	555
17.3.2	Nonlinear Principal Components Analysis	558
17.4	Determining Attribute Weights using Clickstream Analysis	559
17.4.1	Poisson Regression Model	560
17.4.2	Handling Missing Values	560
17.4.3	Choosing Weights Using Poisson Regression	561
17.4.4	Stepwise Poisson Regression Model	562
17.5	Graphical Shopping Interface	562
17.6	E-Commerce Applications	563
17.6.1	MDS Based Product Catalog Map Using Attribute Weights	564
17.6.2	NL-PCA Based Product Catalog Map	568
17.6.3	Graphical Shopping Interface	570
17.7	Conclusions and Outlook	573
	References	574

Part IV Recommender Systems and Communities

18	Communities, Collaboration, and Recommender Systems in Personalized Web Search	579
	Barry Smyth, Maurice Coyle and Peter Briggs	
18.1	Introduction	579
18.2	A Brief History of Web Search	581
18.3	The Future of Web Search	583
18.3.1	Personalized Web Search	584
18.3.2	Collaborative Information Retrieval	588
18.3.3	Towards Social Search	590

18.4	Case-Study 1 - Community-Based Web Search	591
18.4.1	Repetition and Regularity in Search Communities	592
18.4.2	The Collaborative Web Search System	593
18.4.3	Evaluation	596
18.4.4	Discussion	598
18.5	Case-Study 2 - Web Search. Shared.	598
18.5.1	The HeyStaks System	599
18.5.2	The HeyStaks Recommendation Engine	602
18.5.3	Evaluation	604
18.5.4	Discussion	607
18.6	Conclusions	607
	References	609
19	Social Tagging Recommender Systems	615
	Leandro Balby Marinho, Alexandros Nanopoulos, Lars Schmidt-Thieme, Robert Jäschke, Andreas Hotho, Gerd Stumme and Panagiotis Symeonidis	
19.1	Introduction	616
19.2	Social Tagging Recommenders Systems	617
19.2.1	Folksonomy	618
19.2.2	The Traditional Recommender Systems Paradigm	619
19.2.3	Multi-mode Recommendations	620
19.3	Real World Social Tagging Recommender Systems	621
19.3.1	What are the Challenges?	621
19.3.2	BibSonomy as Study Case	622
19.3.3	Tag Acquisition	624
19.4	Recommendation Algorithms for Social Tagging Systems	626
19.4.1	Collaborative Filtering	626
19.4.2	Recommendation based on Ranking	630
19.4.3	Content-Based Social Tagging RS	634
19.4.4	Evaluation Protocols and Metrics	637
19.5	Comparison of Algorithms	639
19.6	Conclusions and Research Directions	640
	References	642
20	Trust and Recommendations	645
	Patricia Victor, Martine De Cock, and Chris Cornelis	
20.1	Introduction	645
20.2	Computational Trust	647
20.2.1	Trust Representation	648
20.2.2	Trust Computation	650
20.3	Trust-Enhanced Recommender Systems	655
20.3.1	Motivation	656
20.3.2	State of the Art	658
20.3.3	Empirical Comparison	664
20.4	Recent Developments and Open Challenges	670

20.5	Conclusions	672
	References	672
21	Group Recommender Systems: Combining Individual Models	677
	Judith Masthoff	
21.1	Introduction	677
21.2	Usage Scenarios and Classification of Group Recommenders	679
21.2.1	Interactive Television	679
21.2.2	Ambient Intelligence	679
21.2.3	Scenarios Underlying Related Work	680
21.2.4	A Classification of Group Recommenders	681
21.3	Aggregation Strategies	682
21.3.1	Overview of Aggregation Strategies	682
21.3.2	Aggregation Strategies Used in Related Work	683
21.3.3	Which Strategy Performs Best	685
21.4	Impact of Sequence Order	686
21.5	Modelling Affective State	688
21.5.1	Modelling an Individual's Satisfaction on its Own	689
21.5.2	Effects of the Group on an Individual's Satisfaction	690
21.6	Using Affective State inside Aggregation Strategies	691
21.7	Applying Group Recommendation to Individual Users	693
21.7.1	Multiple Criteria	693
21.7.2	Cold-Start Problem	695
21.7.3	Virtual Group Members	697
21.8	Conclusions and Challenges	697
21.8.1	Main Issues Raised	697
21.8.2	Caveat: Group Modelling	698
21.8.3	Challenges	698
	References	701
Part V Advanced Algorithms		
22	Aggregation of Preferences in Recommender Systems	705
	Gleb Beliakov, Tomasa Calvo and Simon James	
22.1	Introduction	705
22.2	Types of Aggregation in Recommender Systems	706
22.2.1	Aggregation of Preferences in CF	708
22.2.2	Aggregation of Features in CB and UB Recommendation	708
22.2.3	Profile Construction for CB, UB	709
22.2.4	Item and User Similarity and Neighborhood Formation	709
22.2.5	Connectives in Case-Based Reasoning for RS	711
22.2.6	Weighted Hybrid Systems	711
22.3	Review of Aggregation Functions	712
22.3.1	Definitions and Properties	712
22.3.2	Aggregation Families	716
22.4	Construction of Aggregation Functions	722

22.4.1	Data Collection and Preprocessing	722
22.4.2	Desired Properties, Semantics and Interpretation	724
22.4.3	Complexity and the Understanding of Function Behavior	725
22.4.4	Weight and Parameter Determination	726
22.5	Sophisticated Aggregation Procedures in Recommender Systems: Tailoring for Specific Applications	726
22.6	Conclusions	731
22.7	Further Reading	732
	References	733
23	Active Learning in Recommender Systems	735
	Neil Rubens, Dain Kaplan, and Masashi Sugiyama	
23.1	Introduction	735
23.1.1	Objectives of Active Learning in Recommender Systems	737
23.1.2	An Illustrative Example	738
23.1.3	Types of Active Learning	739
23.2	Properties of Data Points	740
23.2.1	Other Considerations	741
23.3	Active Learning in Recommender Systems	742
23.3.1	Method Summary Matrix	742
23.4	Active Learning Formulation	742
23.5	Uncertainty-based Active Learning	746
23.5.1	Output Uncertainty	746
23.5.2	Decision Boundary Uncertainty	748
23.5.3	Model Uncertainty	749
23.6	Error-based Active Learning	751
23.6.1	Instance-based Methods	752
23.6.2	Model-based	754
23.7	Ensemble-based Active Learning	756
23.7.1	Models-based	756
23.7.2	Candidates-based	757
23.8	Conversation-based Active Learning	760
23.8.1	Case-based Critique	761
23.8.2	Diversity-based	761
23.8.3	Query Editing-based	762
23.9	Computational Considerations	762
23.10	Discussion	763
	References	764
24	Multi-Criteria Recommender Systems	769
	Gediminas Adomavicius, Nikos Manouselis and YoungOk Kwon	
24.1	Introduction	769
24.2	Recommendation as a Multi-Criteria Decision Making Problem	771
24.2.1	Object of Decision	772
24.2.2	Family of Criteria	773

24.2.3	Global Preference Model	774
24.2.4	Decision Support Process	775
24.3	MCDM Framework for Recommender Systems: Lessons Learned	776
24.4	Multi-Criteria Rating Recommendation	780
24.4.1	Traditional single-rating recommendation problem	781
24.4.2	Extending traditional recommender systems to include multi-criteria ratings	782
24.5	Survey of Algorithms for Multi-Criteria Rating Recommenders ...	783
24.5.1	Engaging Multi-Criteria Ratings during Prediction	784
24.5.2	Engaging Multi-Criteria Ratings during Recommendation	791
24.6	Discussion and Future Work	795
24.7	Conclusions	797
	References	798
25	Robust Collaborative Recommendation	805
	Robin Burke, Michael P. O'Mahony and Neil J. Hurley	
25.1	Introduction	805
25.2	Defining the Problem	807
25.2.1	An Example Attack	809
25.3	Characterising Attacks	810
25.3.1	Basic Attacks	810
25.3.2	Low-knowledge attacks	811
25.3.3	Nuke Attack Models	812
25.3.4	Informed Attack Models	813
25.4	Measuring Robustness	814
25.4.1	Evaluation Metrics	815
25.4.2	Push Attacks	816
25.4.3	Nuke Attacks	818
25.4.4	Informed Attacks	819
25.4.5	Attack impact	820
25.5	Attack Detection	820
25.5.1	Evaluation Metrics	821
25.5.2	Single Profile Detection	822
25.5.3	Group Profile Detection	824
25.5.4	Detection findings	827
25.6	Robust Algorithms	828
25.6.1	Model-based Recommendation	828
25.6.2	Robust Matrix Factorisation (RMF)	829
25.6.3	Other Robust Recommendation Algorithms	830
25.6.4	The Influence Limiter and Trust-based Recommendation	831
25.7	Conclusion	832
	References	833
	Index	837

List of Contributors

Gediminas Adomavicius
Department of Information and Decision Sciences
Carlson School of Management, University of Minnesota, Minneapolis, MN 55455,
USA
e-mail: gedas@umn.edu

Xavier Amatriain
Telefonica Research, Via Augusta, 122, Barcelona 08021, Spain
e-mail: xar@tid.es

Riccardo Bambini
Fastweb, via Francesco Caracciolo 51, Milano, Italy
e-mail: riccardo.bambini@fastweb.it

Gleb Beliakov
School of Information Technology, Deakin University, 221 Burwood Hwy,
Burwood 3125, Australia,
e-mail: gleb@deakin.edu.au

Robert Bell
AT&T Labs – Research
e-mail: rbell@research.att.com

David Bonnefoy
Pearltrees,
e-mail: david.bonnefoy@pearltrees.com

Peter Briggs
CLARITY: Centre for Sensor Web Technologies, School of Computer Science &
Informatics, University College Dublin, Ireland,
e-mail: Peter.Briggs@ucd.ie

Robin Burke
Center for Web Intelligence, School of Computer Science, Telecommunication and

Information Systems, DePaul University, Chicago, Illinois, USA

e-mail: rburke@cs.depaul.edu

Tomasa Calvo

Departamento de Ciencias de la Computación, Universidad de Alcalá

28871-Alcalá de Henares (Madrid), Spain.

e-mail: tomasa.calvo@uah.es

Li Chen

Human Computer Interaction Group, School of Computer and Communication Sciences,

Swiss Federal Institute of Technology in Lausanne (EPFL), CH-1015, Lausanne, Switzerland

e-mail: li.chen@epfl.ch

Martine De Cock

Institute of Technology, University of Washington Tacoma, 1900 Pacific Ave, Tacoma, WA, USA (on leave from Ghent University)

e-mail: mdecock@u.washington.edu

Chris Cornelis

Dept. of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281 (S9), 9000 Gent, Belgium

e-mail: Patricia.Victor@ugent.be

Maurice Coyle

CLARITY: Centre for Sensor Web Technologies, School of Computer Science & Informatics, University College Dublin, Ireland,

e-mail: Maurice.Coyle@ucd.ie

Paolo Cremonesi

Politecnico di Milano, p.zza Leonardo da Vinci 32, Milano, Italy Neptuny, via Durando 10, Milano, Italy

e-mail: paolo.cremonesi@polimi.it

Christian Desrosiers

Department of Software Engineering and IT, École de Technologie Supérieure, Montreal, Canada

e-mail: christian.desrosiers@etsmtl.ca

Hendrik Drachsler

Centre for Learning Sciences and Technologies (CELSTEC), Open Universiteit Nederland

e-mail: hendrik.drachsler@ou.nl

Boi Faltings

Artificial Intelligence Laboratory, School of Computer and Communication Sciences

Swiss Federal Institute of Technology in Lausanne (EPFL), CH-1015, Lausanne, Switzerland

e-mail: boi.faltings@epfl.ch

Alexander Felfernig
Graz University of Technology
e-mail: alexander.felfernig@ist.tugraz.at

Gerhard Friedrich
University Klagenfurt
e-mail: gerhard.friedrich@uni-klu.ac.at

Marco de Gemmis
Department of Computer Science, University of Bari “Aldo Moro”, Via E. Orabona,
4, Bari (Italy)
e-mail: degemmis@di.uniba.it

Ulrike Gretzel
Texas A&M University, 2261 TAMU, College Station, TX, USA,
e-mail: ugretzel@tamu.edu

Patrick J.F. Groenen
Econometric Institute, Erasmus University Rotterdam, The Netherlands,
e-mail: groenen@ese.eur.nl

Asela Gunawardana
Microsoft Research, One Microsoft Way, Redmond, WA,
e-mail: aselag@microsoft.com

Andreas Hotho
Knowledge & Data Engineering Group (KDE), University of Kassel, Wilhelmshö,
her Allee 73, 34121 Kassel, Germany,
e-mail: hotho@cs.uni-kassel.de

Hans Hummel
Centre for Learning Sciences and Technologies (CELSTEC), Open Universiteit
Nederland
e-mail: hans.hummel@ou.nl

Neil J. Hurley
School of Computer Science and Informatics, University College Dublin, Ireland
e-mail: neil.hurley@ucd.ie

Robert Jäschke
Knowledge & Data Engineering Group (KDE), University of Kassel, Wilhelmshö
her Allee 73, 34121 Kassel, Germany,
e-mail: jaeschke@cs.uni-kassel.de

Alejandro Jaimes
Yahoo! Research, Av.Diagonal, 177, Barcelona 08018, Spain
e-mail: ajaimes@yahoo-inc.com

Simon James
School of Information Technology, Deakin University, 221 Burwood Hwy,
Burwood 3125, Australia,
e-mail: sjames@deakin.edu.au

Dietmar Jannach
TU Dortmund
e-mail: dietmar.jannach@tu-dortmund.de

Martijn Kagie
Econometric Institute, Erasmus University Rotterdam, The Netherlands,
e-mail: martijn@kagie.net

Dain Kaplan
Tokyo Institute of Technology, Tokyo, Japan
e-mail: dain@cl.cs.titech.ac.jp

George Karypis
Department of Computer Science & Engineering, University of Minnesota,
Minneapolis, USA
e-mail: karypis@cs.umn.edu

Rob Koper
Centre for Learning Sciences and Technologies (CELSTEC), Open Universiteit
Nederland
e-mail: rob.koper@ou.nl

Yehuda Koren
Yahoo! Research,
e-mail: yehuda@yahoo-inc.com

YoungOk Kwon
Department of Information and Decision Sciences
Carlson School of Management, University of Minnesota, Minneapolis, MN 55455,
USA
e-mail: kwonx052@umn.edu

Pasquale Lops
Department of Computer Science, University of Bari “Aldo Moro”, Via E. Orabona,
4, Bari (Italy)
e-mail: lops@di.uniba.it

Nikos Manouselis
Greek Research and Technology Network (GRNET S.A.)
56 Messogeion Av., 115 27, Athens, Greece
e-mail: nikosm@grnet.gr

Leandro Balby Marinho
Information Systems and Machine Learning Lab (ISMLL), University of
Hildesheim, Marienburger Platz 22, 31141 Hildesheim, Germany,
e-mail: marinho@ismll.uni-hildesheim.de

Judith Masthoff

University of Aberdeen, AB24 3UE Aberdeen UK,

e-mail: j.masthoff@abdn.ac.uk

Lorraine McGinty

UCD School of Computer Science and Informatics, University College Dublin,
Dublin 4, Ireland.

e-mail: lorraine.mcginty@ucd.ie

Kevin Mercer

Loughborough University,

e-mail: K.C.Mercer@lboro.ac.uk

Alexandros Nanopoulos

Information Systems and Machine Learning Lab (ISMLL), University of
Hildesheim, Marienburger Platz 22, 31141 Hildesheim, Germany,

e-mail: nanopoulos@ismll.uni-hildesheim.de

Michael P. O'Mahony

CLARITY: Centre for Sensor Web Technologies, School of Computer Science and
Informatics, University College Dublin, Ireland

e-mail: michael.p.omahony@ucd.ie

Nuria Oliver

Telefonica Research, Via Augusta, 122, Barcelona 08021, Spain

e-mail: nuriao@tid.es

Jérôme Picault

Alcatel-Lucent Bell Labs,

e-mail: jerome.picault@alcatel-lucent.com

Pearl Pu

Human Computer Interaction Group, School of Computer and Communication
Sciences,

Swiss Federal Institute of Technology in Lausanne (EPFL), CH-1015, Lausanne,
Switzerland

e-mail: pearl.pu, li.chen, jiyong.zhang@epfl.ch

Josep M. Pujol

Telefonica Research, Via Augusta, 122, Barcelona 08021, Spain

e-mail: jmps@tid.es

Maryam Ramezani

Center for Web Intelligence, College of Computing and Digital Media, 243 S.
Wabash Ave., DePaul University, Chicago, Illinois, USA

e-mail: mramezani@depaul.edu

James Reilly

Google Inc., 5 Cambridge Center, Cambridge, MA 02142, United States.

e-mail: jamesreilly@google.com

Myriam Ribière
Alcatel-Lucent Bell Labs,
e-mail: myriam.ribiere@alcatel-lucent.com

Francesco Ricci
Faculty of Computer Science, Free University of Bozen-Bolzano, Italy
e-mail: fricci@unibz.it

Lior Rokach
Department of Information Systems Engineering, Ben-Gurion University of the
Negev, Israel
e-mail: liorrk@bgu.ac.il

Neil Rubens
University of Electro-Communications, Tokyo, Japan,
e-mail: rubens@hrstc.org

Lars Schmidt-Thieme
Information Systems and Machine Learning Lab (ISMLL), University of
Hildesheim, Marienburger Platz 22, 31141 Hildesheim, Germany,
e-mail: schmidt-thieme@ismll.uni-hildesheim.de

Giovanni Semeraro
Department of Computer Science, University of Bari “Aldo Moro”, Via E. Orabona,
4, Bari (Italy)
e-mail: semeraro@di.uniba.it

Guy Shani
Department of Information Systems Engineering, Ben-Gurion University of the
Negev, Beer-Sheva, Israel
e-mail: shanigu@bgu.ac.il

Bracha Shapira
Department of Information Systems Engineering, Ben-Gurion University of the
Negev, Israel
e-mail: bshapira@bgu.ac.il

Barry Smyth
CLARITY: Centre for Sensor Web Technologies, School of Computer Science &
Informatics, University College Dublin, Ireland,
e-mail: Barry.Smyth@ucd.ie

Gerd Stumme
Knowledge & Data Engineering Group (KDE), University of Kassel, Wilhelmshö-
her Allee 73, 34121 Kassel, Germany,
e-mail: stumme@cs.uni-kassel.de

Masashi Sugiyama
Tokyo Institute of Technology, Tokyo, Japan
e-mail: sugi@cs.titech.ac.jp

Panagiotis Symeonidis

Department of Informatics, Aristotle University, 54124 Thessaloniki, Greece,

e-mail: symeon@csd.auth.gr

Nava Tintarev

University of Aberdeen, Aberdeen, U.K,

e-mail: n.tintarev@abdn.ac.uk

Roberto Turrin

Politecnico di Milano, p.zza Leonardo da Vinci 32, Milano, Italy Neptun, via Durando 10, Milano, Italy

e-mail: roberto.turrin@polimi.it

Alexander Tuzhilin

Department of Information, Operations and Management Sciences

Stern School of Business, New York University

e-mail: atuzhili@stern.nyu.edu

Paolo Viappiani

Department of Computer Science, University of Toronto, 6 King's College Road, M5S3G4, Toronto, ON, CANADA

e-mail: paolo.viappiani@gmail.com

Patricia Victor

Dept. of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281 (S9), 9000 Gent, Belgium

e-mail: Chris.Cornelis@ugent.be

Riina Vuorikari

European Schoolnet (EUN), 24, Rue Paul Emile Janson, 1050 Brussels, Belgium

e-mail: riina.vuorikari@eun.org

Michiel van Wezel

Econometric Institute, Erasmus University Rotterdam, The Netherlands,

e-mail: mvanwezel@acm.org

Kyung-Hyan Yoo

William Paterson University, Communication Department, 300 Pompton Road, Wayne, NJ, USA,

e-mail: toinette75@gmail.com

Markus Zanker

University Klagenfurt

e-mail: markus.zanker@uni-klu.ac.at

Jiyong Zhang

Human Computer Interaction Group, School of Computer and Communication Sciences,

Swiss Federal Institute of Technology in Lausanne (EPFL), CH-1015, Lausanne, Switzerland

e-mail: jiyong.zhang@epfl.ch

Chapter 1

Introduction to Recommender Systems Handbook

Francesco Ricci, Lior Rokach and Bracha Shapira

Abstract Recommender Systems (RSs) are software tools and techniques providing suggestions for items to be of use to a user. In this introductory chapter we briefly discuss basic RS ideas and concepts. Our main goal is to delineate, in a coherent and structured way, the chapters included in this handbook and to help the reader navigate the extremely rich and detailed content that the handbook offers.

1.1 Introduction

Recommender Systems (RSs) are software tools and techniques providing suggestions for items to be of use to a user [60, 85, 25]. The suggestions relate to various decision-making processes, such as what items to buy, what music to listen to, or what online news to read.

“Item” is the general term used to denote what the system recommends to users. A RS normally focuses on a specific type of item (e.g., CDs, or news) and accordingly its design, its graphical user interface, and the core recommendation technique used to generate the recommendations are all customized to provide useful and effective suggestions for that specific type of item.

RSs are primarily directed towards individuals who lack sufficient personal experience or competence to evaluate the potentially overwhelming number of alter-

Francesco Ricci

Faculty of Computer Science, Free University of Bozen-Bolzano, Italy e-mail: fricci@unibz.it

Lior Rokach

Department of Information Systems Engineering, Ben-Gurion University of the Negev, Israel e-mail: liorrk@bgu.ac.il

Bracha Shapira

Department of Information Systems Engineering, Ben-Gurion University of the Negev, Israel e-mail: bshapira@bgu.ac.il

native items that a Web site, for example, may offer [85]. A case in point is a book recommender system that assists users to select a book to read. In the popular Web site, Amazon.com, the site employs a RS to personalize the online store for each customer [47]. Since recommendations are usually personalized, different users or user groups receive diverse suggestions. In addition there are also non-personalized recommendations. These are much simpler to generate and are normally featured in magazines or newspapers. Typical examples include the top ten selections of books, CDs etc. While they may be useful and effective in certain situations, these types of non-personalized recommendations are not typically addressed by RS research.

In their simplest form, personalized recommendations are offered as ranked lists of items. In performing this ranking, RSs try to predict what the most suitable products or services are, based on the user's preferences and constraints. In order to complete such a computational task, RSs collect from users their preferences, which are either explicitly expressed, e.g., as ratings for products, or are inferred by interpreting user actions. For instance, a RS may consider the navigation to a particular product page as an implicit sign of preference for the items shown on that page.

RSs development initiated from a rather simple observation: individuals often rely on recommendations provided by others in making routine, daily decisions [60, 70]. For example it is common to rely on what one's peers recommend when selecting a book to read; employers count on recommendation letters in their recruiting decisions; and when selecting a movie to watch, individuals tend to read and rely on the movie reviews that a film critic has written and which appear in the newspaper they read.

In seeking to mimic this behavior, the first RSs applied algorithms to leverage recommendations produced by a community of users to deliver recommendations to an active user, i.e., a user looking for suggestions. The recommendations were for items that similar users (those with similar tastes) had liked. This approach is termed collaborative-filtering and its rationale is that if the active user agreed in the past with some users, then the other recommendations coming from these similar users should be relevant as well and of interest to the active user.

As e-commerce Web sites began to develop, a pressing need emerged for providing recommendations derived from filtering the whole range of available alternatives. Users were finding it very difficult to arrive at the most appropriate choices from the immense variety of items (products and services) that these Web sites were offering.

The explosive growth and variety of information available on the Web and the rapid introduction of new e-business services (buying products, product comparison, auction, etc.) frequently overwhelmed users, leading them to make poor decisions. The availability of choices, instead of producing a benefit, started to decrease users' well-being. It was understood that while choice is good, more choice is not always better. Indeed, choice, with its implications of freedom, autonomy, and self-determination can become excessive, creating a sense that freedom may come to be regarded as a kind of misery-inducing tyranny [96].

RSs have proved in recent years to be a valuable means for coping with the information overload problem. Ultimately a RS addresses this phenomenon by pointing

a user towards new, not-yet-experienced items that may be relevant to the users current task. Upon a user's request, which can be articulated, depending on the recommendation approach, by the user's context and need, RSs generate recommendations using various types of knowledge and data about users, the available items, and previous transactions stored in customized databases. The user can then browse the recommendations. She may accept them or not and may provide, immediately or at a next stage, an implicit or explicit feedback. All these user actions and feedbacks can be stored in the recommender database and may be used for generating new recommendations in the next user-system interactions.

As noted above, the study of recommender systems is relatively new compared to research into other classical information system tools and techniques (e.g., databases or search engines). Recommender systems emerged as an independent research area in the mid-1990s [35, 60, 70, 7]. In recent years, the interest in recommender systems has dramatically increased, as the following facts indicate:

1. Recommender systems play an important role in such highly rated Internet sites as Amazon.com, YouTube, Netflix, Yahoo, Tripadvisor, Last.fm, and IMDb. Moreover many media companies are now developing and deploying RSs as part of the services they provide to their subscribers. For example Netflix, the online movie rental service, awarded a million dollar prize to the team that first succeeded in improving substantially the performance of its recommender system [54].
2. There are dedicated conferences and workshops related to the field. We refer specifically to ACM Recommender Systems (RecSys), established in 2007 and now the premier annual event in recommender technology research and applications. In addition, sessions dedicated to RSs are frequently included in the more traditional conferences in the area of data bases, information systems and adaptive systems. Among these conferences are worth mentioning ACM SIGIR Special Interest Group on Information Retrieval (SIGIR), User Modeling, Adaptation and Personalization (UMAP), and ACM's Special Interest Group on Management Of Data (SIGMOD).
3. At institutions of higher education around the world, undergraduate and graduate courses are now dedicated entirely to RSs; tutorials on RSs are very popular at computer science conferences; and recently a book introducing RSs techniques was published [48].
4. There have been several special issues in academic journals covering research and developments in the RS field. Among the journals that have dedicated issues to RS are: AI Communications (2008); IEEE Intelligent Systems (2007); International Journal of Electronic Commerce (2006); International Journal of Computer Science and Applications (2006); ACM Transactions on Computer-Human Interaction (2005); and ACM Transactions on Information Systems (2004).

In this introductory chapter we briefly discuss basic RS ideas and concepts. Our main goal is not much to present a self-contained comprehensive introduction or survey on RSs but rather to delineate, in a coherent and structured way, the chapters

included in this handbook and to help the reader navigate the extremely rich and detailed content that the handbook offers.

The handbook is divided into five sections: techniques; applications and evaluation of RSs; interacting with RSs; RSs and communities; and advanced algorithms.

The first section presents the techniques most popularly used today for building RSs, such as collaborative filtering; content-based, data mining methods; and context-aware methods.

The second section surveys techniques and approaches that have been utilized to evaluate the quality of the recommendations. It also deals with the practical aspects of designing recommender systems; describes design and implementation considerations; and sets guidelines for selecting the more suitable algorithms. The section also considers aspects that may affect RS design (domain, device, users, etc.). Finally, it discusses methods, challenges and measures to be applied in evaluating the developed systems.

The third section includes papers dealing with a number of issues related to how recommendations are presented, browsed, explained and visualized. The techniques that make the recommendation process more structured and conversational are discussed here.

The fourth section is fully dedicated to a rather new topic, exploiting user-generated content (UGC) of various types (tags, search queries, trust evaluations, etc.) to generate innovative types of recommendations and more credible ones. Despite its relative newness, this topic is essentially rooted in the core idea of a collaborative recommender,

The last selection presents papers on various advanced topics, such as: the exploitation of active learning principles to guide the acquisition of new knowledge; suitable techniques for protecting a recommender system against attacks of malicious users; and RSs that aggregate multiple types of user feedbacks and preferences to build more reliable recommendations.

1.2 Recommender Systems Function

In the previous section we defined RSs as software tools and techniques providing users with suggestions for items a user may wish to utilize. Now we want to refine this definition illustrating a range of possible roles that a RS can play. First of all, we must distinguish between the role played by the RS on behalf of the service provider from that of the user of the RS. For instance, a travel recommender system is typically introduced by a travel intermediary (e.g., Expedia.com) or a destination management organization (e.g., Visitfinland.com) to increase its turnover (Expedia), i.e., sell more hotel rooms, or to increase the number of tourists to the destination [86]. Whereas, the user's primary motivations for accessing the two systems is to find a suitable hotel and interesting events/attractions when visiting a destination.

In fact, there are various reasons as to why service providers may want to exploit this technology: