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

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
May 2010 *Paul B. Kantor*

Contents

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 Neptuny, 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`

F. Ricci et al. (eds.), *Recommender Systems Handbook*,
DOI 10.1007/978-0-387-85820-3_1,

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: