Statistical Methods for Recommender Systems

Designing algorithms to recommend items such as news articles and movies to users is a challenging task in numerous web applications. The crux of the problem is to rank items based on past user responses to optimize for multiple objectives. Major technical challenges are high-dimensional prediction with sparse data and constructing high-dimensional sequential designs to collect data for user modeling and system design.

This comprehensive treatment of the statistical issues that arise in recommender systems includes detailed, in-depth discussions of current state-of-the-art methods such as adaptive sequential designs (multiarmed bandit methods), bilinear random-effects models (matrix factorization), and scalable model fitting using modern computing paradigms such as MapReduce. The authors draw on their vast experience working with such large-scale systems at Yahoo! and LinkedIn and bridge the gap between theory and practice by illustrating complex concepts with examples from applications with which they are directly involved.

DR. DEEPAK K. AGARWAL is a big data analyst with several years of experience developing and deploying state-of-the-art machine learning and statistical methods for improving the relevance of web applications. He is also experienced in conducting new scientific research to solve difficult big data problems, especially in the areas of recommender systems and computational advertising. He is a Fellow of the American Statistical Association and associate editor of top-tier journals in statistics.

DR. BEE-CHUNG CHEN is a leading technologist with extensive industrial and research experience in developing state-of-the-art recommender systems. He has been a key designer of the recommendation algorithms that power the LinkedIn home page and mobile feeds, the Yahoo! home page, Yahoo! News, and other sites. His research areas include recommender systems, data mining, machine learning, and big data analytics.

For Bharati Agarwal and Shiao-Ching Chung

Cambridge University Press
978-1-107-03607-9 - Statistical Methods for Recommender Systems
Deepak K. Agarwal and Bee-chung Chen
Frontmatter
More information

Statistical Methods for Recommender Systems

DEEPAK K. AGARWAL
LinkedIn Corporation

BEE-CHUNG CHEN
LinkedIn Corporation

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107036079

First published 2016

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data
Agarwal, Deepak K., 1973– author.
Statistical methods for recommender systems / Deepak K. Agarwal, Yahoo! Research, Bee Chung-Chen, Yahoo! Research.
pages cm
ISBN 978-1-107-03607-9
1. Recommender systems (Information filtering) – Statistical methods. 2. Expert systems (Computer science) – Statistical methods. I. Chung-Chen, Bee, author. II. Title.
QA76.76.E95A395 2016
006.3′3–dc23 2015026092

ISBN 978-1-107-03607-9 Hardback

Additional resources for this publication at
https://github.com/beechung/Latent-Factor-Models

Contents

Preface

What This Book Is About

Recommender systems are automated computer programs that match items to users in different contexts. Such systems are ubiquitous and have become an integral part of our daily lives. Examples include recommending products to users on a site like Amazon, recommending content to users visiting a website like Yahoo!, recommending movies to users on a site like Netflix, recommending jobs to users on a site like LinkedIn, and so on. The matching algorithms are constructed using large amounts of high-frequency data obtained from past user interactions with items. The algorithms are statistical in nature and involve challenges in areas like sequential decision processes, modeling interactions with very high-dimensional categorical data, and developing scalable statistical methods. New methodologies in this area require close collaboration among computer scientists, machine learners, statisticians, optimization experts, system experts, and, of course, domain experts. It is one of the most exciting applications of big data.

Why We Wrote This Book

Although much has been written about recommender systems in various fields, such as computer science, machine learning, and statistics, focusing on specific aspects of the problem, a comprehensive treatment of all statistical issues and how they are interrelated is lacking. We came to this realization while deploying such systems at Yahoo! and LinkedIn. For instance, much of the focus in statistics and machine learning is on building models that minimize out-of-sample predictive error. However, this does not address all aspects of practical importance. Statistically, a recommender system is a high-dimensional sequential process, and it is equally important to study issues like design of experiments as it is to develop sophisticated statistical models. In fact, the two are closely related – efficient design needs models to tame the curse of dimensionality. Also, most existing work in the literature tends to build models for univariate response, such as movie ratings, purchases, and click rates. With the advent of social media outlets like Facebook, LinkedIn, and Twitter, multiple responses are available. For instance, one may want to model click rates, share rates, and tweet rates simultaneously for a news recommender application. Such multivariate response models are challenging to build. Finally, given the machinery to obtain such multivariate predictions, how does one construct utility functions to make recommendations? Is it more important to optimize share rates relative to click rates? Answers to these types of questions can be obtained through multiobjective optimization working in close collaboration with domain experts to elicit some utility parameters.

The goal of this book is to provide a comprehensive discussion of all such issues that arise in the context of recommender systems. This is in addition to a detailed and in-depth discussion of current state-of-the-art statistical methods that include techniques like adaptive sequential designs (multiarmed bandit methods), bilinear random-effects models (matrix factorization), and scalable model fitting using modern-distributed computing infrastructure. Our goal in writing this book is to draw on our vast experience working with such large-scale systems in industrial settings and to bring these issues to the attention of the statistical, machine learning, and computer science communities. We believe this will be beneficial in a number of ways. It may help in advancing methodological research in high-dimensional and big data statistics, especially for web applications. We understand that conducting such research in an academic setting requires access to software that can run on massive data. To facilitate this, we supplement the book with open source software: https://github.com/beechung/Latent-Factor-Models. We also believe the book will help in bridging the gap between theory and applications. It will provide problem owners with a good understanding of the statistical issues involved and modelers with an in-depth understanding of statistical issues that arise in practical applications that are rather complex.

Organization

We divide the content of the book into three parts.

In Part I, we introduce the recommender system problem, challenges in the problem, main ideas used to tackle the challenges, and the required background knowledge. In Chapter 2, we give an overview of classical methods that have been used to develop recommender systems. Such methods involve characterizing users and items as feature vectors and then scoring user-item pairs based on some similarity function, standard supervised learning, or collaborative filtering. These classical methods usually ignore the explore-exploit trade-off in recommender problems. Hence, in Chapter 3, we discuss the importance of this issue and introduce the main ideas that will be used to solve the issue in later chapters. Before we delve into technical solutions, in Chapter 4, we review a variety of methods for evaluating the performance of different recommendation algorithms.

In Part II, we provide detailed solutions to common problem settings. We start with an introduction to various problem settings and an example system architecture in Chapter 5, and then we devote the next three chapters to three common problem settings. Chapter 6 provides solutions to the most-popular recommendation problem, with a special focus on the explore-exploit aspect. Chapter 7 deals with personalized recommendation through feature-based regression, with an emphasis on how to continuously update the model(s) to leverage the most recent user-item interaction data and quickly converge to a good solution. Chapter 8 extends the methods developed in Chapter 7 from feature-based regression to factor models (matrix factorization) and, at the same time, provides a natural solution to the cold-start problem in factor models.

In Part III, we present three advanced topics. In Chapter 9, we present a factorization model that simultaneously identifies topics in items and users' affinities with different topics through a modified matrix factorization model that uses the latent Dirichlet allocation (LDA) topic model. In Chapter 10, we investigate context-dependent recommender problems, in which the recommended items not only need to have high affinity with the user but also have to be relevant to the context (e.g., recommending items related to a news article that the user is currently reading). In Chapter 11, we discuss a principled framework for optimizing multiple objectives based on a constrained optimization approach, where we seek to maximize one objective (e.g., revenue) subject to bounded loss in other objectives (e.g., no more than 5 percent loss in clicks).

Limitations

Like all books, ours has limitations. We do not provide an in-depth coverage of modern computational paradigms, such as Spark, that can be used to fit some of the models presented at scale. Online evaluation of models when users form a social graph cannot be done properly with traditional experimental design methods. New techniques that can adjust for interference because of social graphs need to be developed. We do not cover such advanced topics in this book. Throughout, we address the problem of recommendations through a response prediction approach using regression as our main tool. This is primarily because we believe that output from these models is easy to combine with downstream utilities. We do not provide a comprehensive coverage of methods that are based on direct optimization of ranking loss functions. A comparison of the two approaches would also be a worthwhile topic for discussion.

Acknowledgement

Our special thanks to Raghu Ramakrishnan, Liang Zhang, Xuanhui Wang, Pradheep Elango, Bo Long, Bo Pang, Rajiv Khanna, Nitin Motgi, Seung-Taek Park, Scott Roy, Joe Zachariah for many insightful discussions and collaboration. We would also like to thank our colleagues both at Yahoo! and LinkedIn for all the encouragement and support without which many of the ideas we had would not see the light of the day.