Advanced Database Marketing

Innovative Methodologies and Applications for Managing Customer Relationships

EDITED BY
Kristof Coussement
Koen W. De Bock
Scott A. Neslin

ROUTLEDGE

A Gower Book

To my lovely wife Ilse for her endless support.
To my little funny rascals.
To my dear parents Carline and Wim.
Kristof Coussement

To my prime source of reflection, fun, support and love, Jessie.
To my dear brother Steven and my parents Roger and Patricia.
Koen W. De Bock

To my wife Betsey for her support, wisdom, and good cheer.
To our children Jenny, Louise, and Lenny for their inspiration and joie de vivre.
To my parents Lenora and Stanley who gave me everything, and more.
Scott A. Neslin

Advanced Database Marketing

Innovative Methodologies and Applications for Managing Customer Relationships

Edited by
KRISTOF COUSSEMENT
KOEN W. DE BOCK
and
SCOTT A. NESLIN

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2013 by Gower Publishing

Published 2016 by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN
711 Third Avenue, New York, NY 10017, USA

Routledge is an imprint of the Taylor & Francis Group, an informa business

Gower Applied Business Research
Our programme provides leaders, practitioners, scholars and researchers with thought provoking, cutting edge books that combine conceptual insights, interdisciplinary rigour and practical relevance in key areas of business and management.

British Library Cataloguing in Publication Data
Advanced database marketing : innovative methodologies and applications for managing customer relationships.
1. Database marketing. 2. Consumer profiling.
I. Coussement, Kristof. II. Bock, Koen W. de. III. Neslin, Scott A., 1952-
658.8'34-dc23

Library of Congress Cataloging-in-Publication Data
Advanced database marketing : innovative methodologies and applications for managing customer relationships / [edited] by Kristof Coussement, Koen W. De Bock, and Scott A. Neslin.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-4094-4461-9 (hbk)
1. Database marketing. 2. Consumer profiling. 3. Customer relations. I. Coussement, Kristof. II. Bock, Koen W. de. III. Neslin, Scott A., 1952-
HF5415.126.A377 2013
658.8'120285574--dc23

2012032138

ISBN 978 1 4094 4461 9 (hbk)
ISBN 978 1 3155 6568 2 (ebk)

Contents

List of Figures

List of Tables

About the Editors

Kristof Coussement is Associate Professor of Marketing at IÉSEG School of Management (LEM-CNRS) of the Catholic University of Lille in France and co-director of the Expertise Center for Database Marketing. He has served as a Visiting Scholar at EMLyon Business School (2011).

Professor Coussement teaches several marketing-related courses including *Customer Relationship Management*, *Database Marketing* and *Strategic Marketing Research* in which students are taught the theoretical principles of all aspects in operational and analytical customer relationship management (CRM), the methodological foundations of predictive marketing modeling, and marketing research.

Professor's Coussement's main research interests are all aspects in customer intelligence, B-to-B (business to business) intelligence, direct marketing, and analytical CRM. Improving his "practical" experience over the years by doing several real-life research projects in a different number of industries, his main focus lies on doing profound academic research with a high added value to business. He publishes in international peer-reviewed journals such as *Decision Support Systems*, *Information & Management*, *European Journal of Operational Research*, *European Journal of Marketing*, *Journal of Business Research*, *Computational Statistics & Data Analysis* and *Expert Systems with Applications*. These works have been presented at various conferences around the world. Moreover, he is co-author of the book *Marketing Research With SAS Enterprise Guide*.

Professor Coussement is founder and committee member of BAQMaR, the largest online European Association for Quantitative & Qualitative Marketing Research. More information about his work can be found at www.kristofcoussement.com

Koen W. De Bock is Assistant Professor of Marketing at the IÉSEG School of Management (LEM-CNRS) with campuses in Lille and Paris in France. He is a co-director of the Expertise Center for Database Marketing. In November 2011, Professor De Bock taught at the MBA program at the University of Stellenbosch Business School (USB), Cape Town, South Africa as visiting faculty.

In the Master in Management, he teaches the Internet marketing-related courses Digital Marketing, Web Analytics, Web Advertising, Social Media Marketing, and Search Engine Marketing. Moreover, he conducts research in database marketing and Internet marketing and is especially passionate about big-data machine learning methodologies for predictive modeling in marketing that generate true business value. He has published in international journals such as

Journal of Business Research, *Computational Statistics & Data Analysis*, and *Expert Systems with Applications* and has presented his work at various international academic and industrial conferences. More information can be found at www.kdnuggets.com/koendebock.be

Scott A. Neslin is Albert Wesley Frey Professor of Marketing at the Tuck School of Business, Dartmouth College. He has been at the Tuck School since completing his PhD in 1978 at the Sloan School of Management, MIT. He has ser ved as a Visiting Scholar at the School of Management, Yale University (1989–1990), the Teradata/Duke CRM Center located at the Fuqua School of Management, Duke University (2002), and the Graduate School of Business, Columbia University (2009–2010). Professor Neslin' s expertise includes database marketing, sales promotion, and marketing mix models. He has researched the application of predictive modeling for cross-selling and identifying customer "churners," evaluated the effectiveness of loyalty programs, and examined the determinants of customer channel migration and "research shopping." He has investigated the impact of promotion on consumer stockpiling, brand loyalty, and consumption, as well as the role of advertising in reinforcing brand loyalty. He has published several articles on these topics in journals such as *Marketing Science*, *Journal of Marketing Research*, *Journal of Marketing*, and *Journal of Interactive Marketing*. He is co-author, with Professor Robert C. Blattberg, of Sales *Promotion: Concepts, Methods, and Strategies*; co-author, with Robert C. Blattberg and Byung-Do Kim, of *Database Marketing: Analyzing and Managing Customers*, and author of the monograph *Sales Promotion*. He is Associate Editor of *Marketing Science*, and on the editorial boards of the *Journal of Marketing Research*, *Journal of Marketing*, *Journal of Interactive Marketing*, *Journal of the Academy of Marketing Science*, and *Marketing Letters*. He serves as Senior Advisor for Resource Systems Group, Inc.

List of Contributors

Asim Ansari, Columbia Business School, Columbia University, New York, USA

Bart Baesens, Katholieke Universiteit Leuven, Belgium

Dries F. Benoit, Universiteit Gent, Belgium

Randolph E. Bucklin, Anderson School of Management, University of California, Los Angeles, USA

Kristof Coussement, IÉSEG School of Management, Université Catholique de Lille (LEM, UMR CNRS 8179), France

Koen W. De Bock, IÉSEG School of Management, Université Catholique de Lille (LEM, UMR CNRS 8179), France

Anindya Ghose, New York University's Leonard N. Stern School of Business, USA

Yuan Yuan Guo, Lingnan University, Tuen Mun, Hong Kong

Sang-Pil Han, City University of Hong Kong, College of Business, Hong Kong

Stefan Lessmann, Universität Hamburg, Germany

David Martens, Universiteit Antwerpen, Belgium

Dina Mayzlin, Yale School of Management, USA

Wendy W. Moe, Robert H. Smith School of Business, University of Mar yland, USA

Scott A. Neslin, Tuck School of Business, Dartmouth College, USA

Oliver J. Rutz, Michael G. Foster School of Business at the University of Washington, USA

Dirk Van den Poel, Universiteit Gent, Belgium

Wouter Verbeke, Edinburgh Business School, United Kingdom

Thomas Verbraken, Katholieke Universiteit Leuven, Belgium

Griet Verhaert, Direct Social Communication, Belgium

Véronique Van Vlasselaer, Katholieke Universiteit Leuven, Belgium

Man Leung Wong, Lingnan University, Tuen Mun, Hong Kong

Preface

1 Book Rationale

Marketing analytics – and database marketing in particular – is recognized as an important business discipline, and its popularity and adoption is still growing. According to the 2012 CMO Survey, partnered by Duke University's Fuqua School of Business, the American Marketing Association and McKinsey & Company spending on marketing analytics is expected to increase by 60 percent over the next three years from 5.7 per cent to 9.1 per cent of the marketing budget, and the size of customer analytic teams is expected to grow by nearly 20 percent (The CMO Survey, 2012). Analytical applications and performance management software's revenue reached $12.2 billion in 2011, representing a 16.4 percent increase in comparison to 2010 (Aquino, 2012). Furthermore, database marketing is a well-researched academic discipline. The popularity of the database marketing field in terms of the amount of scholarly journal publications is growing, especially during the last decade.

Moreover, database marketing thrives on the cross-fertilization of ideas and approaches. First, academics and corporations exchange best practices with one another. Academic research in database marketing is strongly business-driven. Many novel methodologies and applications are inspired by actual business problems, and researchers often seek collaborations with businesses which inspire them and give them access to unique customer data. For instance the reputed Customer Analytics Initiative at the Wharton School of the University of Pennsylvania focuses on the development and application of customer analytic methods with the idea to connect leading academic research institutions with the business world. For business professionals, academic research is often seen as the origin of innovative approaches to handle data and analyze customers' behavior in the continuing effort to create competitive advantage. Second, cross-fertilization exists not only between academia and the professional world, but also among various (academic) disciplines. Database marketing draws on the interplay among fields such as marketing, statistics, operations research, management information systems, and computer science.

A crucial recent trend in database marketing is the veritable explosion of techniques, capabilities, and applications. Advances in machine learning, econometrics, and optimization, combined with more powerful computing resources for applying these techniques, have created new opportunities for database marketers. Nowhere is this more apparent than in the field of internet marketing, where digital advertising, recommendation systems, social media, and mobile marketing are in the exponential phases of growth.

The rationale of this book is the following. We want to popularize and review recent advances in the field of database marketing. We see this book as a complement to previously published database marketing books as we assembled chapters that represent the most up-to-date methodologies to analyze customer data and deliver insights into the fastest emerging applications wherein database marketing plays a crucial role.

2 Target Audience

This book is of interest to both academics and professionals by offering an overview of innovative methodologies and applications in the database marketing field. The target audience includes:

- *MSc/Masters students* in marketing analytics;
- *academics* and *PhD students* who are interested in researching and teaching in the field of database marketing; and
- *marketing analysts* in business with at least a medium understanding of the basic principles in database marketing.

The unique selling proposition of this book is to offer the above audiences an integrated set of methods and applications that go a step further than introductory books on database marketing. We hope to offer masters students an appreciation for the capabilities of modern database marketing, triggering their interest in novel areas. Academics and PhD students should consider this a valuable reference work and a cutting-edge overview of what is currently "hot" in academia and business. Finally, we are convinced that due to saturated markets and fierce competition, analysts in business search for innovative ways of creating unique *customer intelligence*, something that is found in this book.

3 Structure of the Book

This book describes recent advances in database marketing. The chapters are organized in terms of Methods and Applications, as shown in Figure P1.1.

Part I describes the recent advances in the *methods* used in the database marketing domain. The structure of Part I follows the database marketing analysis process with:

1. assembling data (Chapters 1–2);
2. modeling the data (Chapters 3–5); and
3. implementing model results (Chapter 6).

Part II illustrates innovative *applications* in the database marketing field. The majority of them are situated in the field of Internet marketing. Chapter 7 explores the field of the recommender systems for generating one-to-one personalized offers. Chapter 8 investigates the new field of mobile marketing, spawned by the increased popularity of smartphones that allow customers to create, interact, and share content. Furthermore, firms are exposed to new opportunities and challenges to optimize their online advertising strategies (Chapters 9–10), while various strategies are discussed in Chapter 11 about how firms can manage online social interactions. Chapter 12 details customer management approaches for maximizing customer lifetime value (CLV) by combining the power of predictive models and dynamic programming, while the last chapter reveals the best practices of applying traditional direct marketing activities in the non-profit sector.

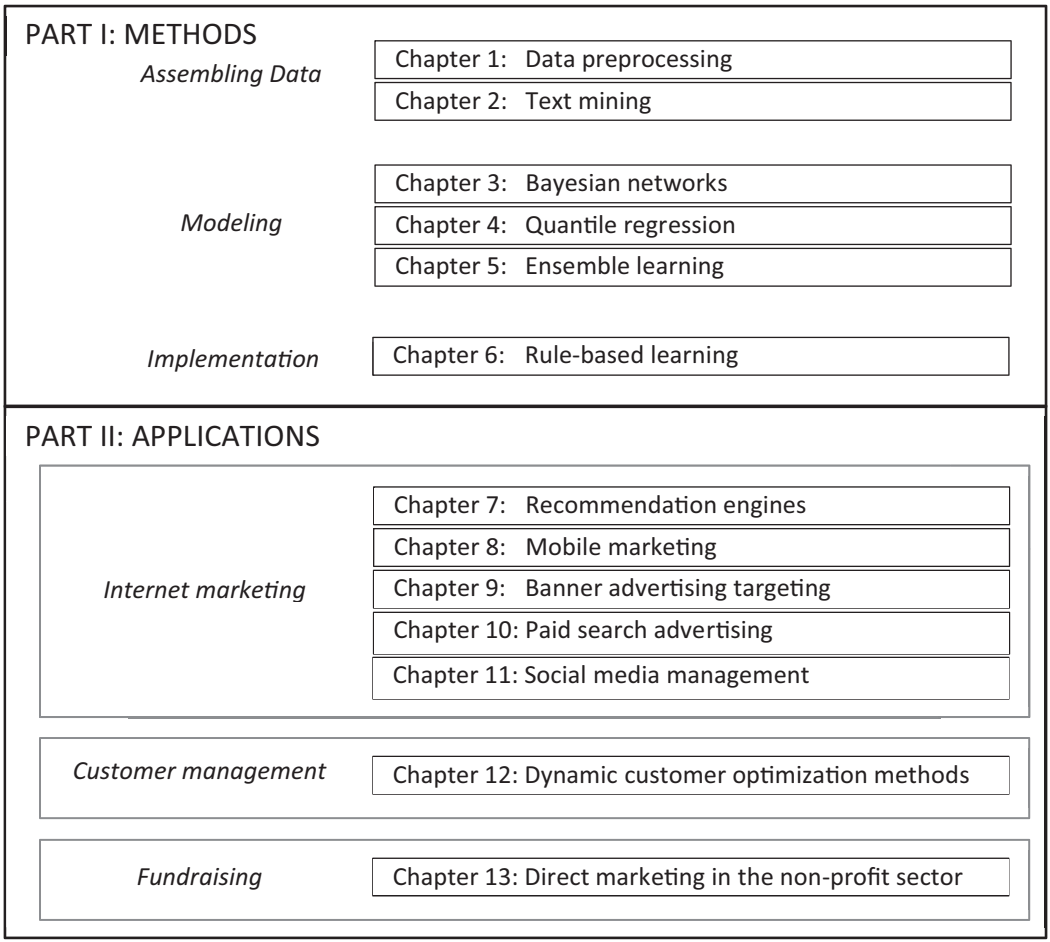

Figure P1.1 Book structure and chapter overview

4 Acknowledgments

First, the editors would like to thank all contributing authors for sharing their expertise. We greatly appreciate the contributors' motivation throughout this book project. Furthermore, we thank all database marketing researchers who built the foundations of and who contributed to our domain. Next, database marketing is a management discipline and its study is irrelevant in the absence of the practical relevance for business. Therefore, we would like to thank all marketing analysts working in firms around the globe who daily inspire our research, and who act as ambassadors of our domain on a day-to-day basis. We thank our home institutions, the IÉSEG School of Management and the Tuck School of Business, Dartmouth College, for providing the resources and research support that make possible an undertaking such as this. Lastly, we extend special thanks to Jonathan Norman and his team at Gower/Ashgate Publishing for well-receiving the book idea and providing us with all the necessary support.

References

Aquino, J. (2012). Business intelligence and analytics are getting hotter . *CRM Magazine*, June 2012, 15.

The CMO Survey (2012). Highlights and Insights Februar y 2012. A vailable at: http://www.cmosurvey.org [accessed October 15, 2012].

Reviews for

Advanced Database Marketing: Innovative Methodologies and Applications for Managing Customer Relationships

This is a great book for masters or doctoral students, academics and practitioners interested in learning the cutting-edge knowledge in database marketing. It provides a comprehensive coverage on the recent advances in the methods and applications in the area. The topics are interesting, relevant and managerially useful.

Shibo Li, Indiana University, USA

This is an excellent contribution to current knowledge on database marketing, an essential dynamic field that demands continuous learning from marketers. The book covers state-of-the-art approaches in classical database marketing areas, yet also in emerging areas such as mobile, social media and Internet advertising. It is highly recommend for quantitative consultants, market researchers and managers as well as academic researchers in this field.

Barak Libai, Arison School of Business, Interdisciplinary Center (IDC), Israel

Database marketing has a long history, but it is rapidly evolving in new directions. This book has all the important new content: Internet and web, mobile, social, etc. It's a fantastic source of up-to-date knowledge for both students and practitioners.

Thomas H. Davenport, Harvard Business School, USA and co-founder
and Director of Research, International Institute for Analytics

In a time where Big Data – the analysis of large datasets – has been frequently named as the main source of competitive advantage in future, this book is essential reading for researchers and managers alike. It provides an excellent in-depth overview of methods and exemplary applications illustrating the potential of generating insights from large-scale databases.

Michael Haenlein, Professor of Marketing, ESCP Europe

WOW! As a 20-year practitioner of database marketing, I found this book to be packed full of practical applications on a wide range of topics within a theoretical framework. An astonishingly rich resource for anyone with intentions to increase lifetime value of the customer, not just measure it.

Peter Liberatore, Senior Manager of Customer Analytics, L.L.Bean

Advanced Database Marketing introduces state-of-the-art methods in marketing and business analytics that firms can use to extract meaningful information from the wide range of available data. Contributions by leading researchers create an easy-to-read collection of chapters that cover data analytics (such as text mining, Bayesian networks, and quartile regression) as well as comprehensive reviews of most timely applications (such as recommendation systems, mobile marketing, online advertising, and online social interaction management). This book is particularly valuable to managers in any firm that uses the Internet

for e-commerce or social media work and/or has access to individual customer-level data and, further, an indispensable asset to academics interested in a comprehensive introduction to the field of database marketing and an inventory of its current state-of-the-art.

Prof. dr. Jacob Goldenberg, Hebrew University of Jerusalem, Israel and
co-editor of the *International Journal of Research in Marketing*

Introduction

The aim of marketing is to know and understand the customer so well the product or service fits him and sells itself.

(Peter F. Drucker, 1973)

1 The Brave New World of Database Marketing

The origins of database marketing can be traced to the fields of direct marketing and relationship marketing. Direct marketing brought to the forefront the importance of customer data, concepts such as recency, frequency, and monetary value (RFM), predictive modeling, and the need for accountability in marketing efforts. Relationship marketing, introduced by Leonard Berry (Berry, 1983), broadened the scope of database marketing to consider the customer relationship, exemplified by concepts such as customer acquisition, retention, and development, and the unifying theme of customer lifetime value. The result of this fusion between direct marketing and relationship marketing is what we know as customer relationship management (CRM). Database marketing can be seen as the analytical side of CRM, and is sometimes called analytical CRM.

Blattberg et al. (2008: 4) considered these developments in developing a definition of database marketing, namely:

Database marketing is the use of customer databases to enhance marketing productivity through more effective acquisition, retention, and development of customers.

This definition indeed captures the importance of analyzing customer data, making the analysis accountable (*enhance marketing productivity*) while focusing on the the importance of the customer relationship (*acquisition, retention, and development*).

This definition still applies today. However, there are four recent trends that substantially amplify this definition. These trends are:

1. the greater variety of data available;
2. the availability of a broader set of methodologies beyond standard statistical tools;
3. the desire to develop not only actions but *insights* from the data; and
4. the ability to implement these targeted actions quickly and in real time.

The proliferation of data available today is due in part to advances in data processing, collection, and management, but perhaps most strikingly due to the diffusion and harnessing of the Internet. The Internet provides traditional transactional data such as customer purchases, but much more, including product search behavior,

the formulation as well as use of recommendations, participation in social media, exposure to advertising, and the modern form of direct mail, namely email. Even more important is that companies are starting to merge these data with offline data, creating a "360-degree view" of the customer. Indeed the proliferation of Internet data is why we find ourselves with five chapters strongly related to database applications involving the Internet.

The "tried and true" methodologies of simple RFM analysis and regression are still widely used in database marketing. But new, more powerful methods are making their way into modern database marketing. The distinction of these methods is they draw on several disciplines, including computer science, operations research, computational linguistics, sociology, economics, as well as statistics. This is why methods such as machine learning, dynamic programming, text mining, Bayesian analysis, and consumer choice models make their way into this book.

One criticism of database marketing is that it has been "black-box;" it prescribes actions that work in the sense of increased response rates and profits, but in today's world of database marketing there is more emphasis on insight. The de-mystification of database marketing has become important for two key reasons. First, as database marketing has become a more significant investment, senior marketing management pays more attention to it, and marketing managers want to understand *why* they are contacting customer A but not customer B, and why they are recommending certain products to certain customers. They want to make sure the activities of the database marketing group are consistent with the positioning and target group strategy of the brand. Second, new tools are becoming available for deriving insights. This is particularly evident in the rules-based learning chapter.

Finally, database marketing's emphasis on implementation and accountability has been enhanced by modern-day capabilities. Companies can now implement the recommendations prescribed by statistical models. They can do so in real time on the Internet. They can conduct field tests of recommendation engines, search advertising copy, and banner advertising. In general, companies' ability to implement and evaluate the actions prescribed by sophisticated models can be tested more easily today than ever before.

In summary, the definition of database marketing hasn't changed – it's still about analyzing customer data and using it to improve marketing productivity by managing the customer relationship. But the *meaning* of the definition has become more vivid and more exciting due to the greater variety of data, the increasingly multidisciplinary analytical "toolkit," the drive for insights in addition to financial performance, and the capability to implement and evaluate more effectively. The reader will see these themes emerge in the chapters we have assembled for this book.

2 Book Contents

The contributions within this book are structured along the two dimensions – methodology and application. Part I describes the methodological advances, while Part II summarizes innovative applications areas in the database marketing field. Below you will find a detailed overview of the different chapters in the book.

2.1 PART I: METHODS

During the last few decades, methods for tracking consumer behavior became more sophisticated, and there has been a move from describing historical customer information to predicting consumers' future behavior. Predictive modeling has established itself as a popular tool in database marketing. The true utility of a predictive model depends on the decisions made in:

1. assembling data;
2. modeling; and
3. implementation.

Part I is structured along these three stages of the predictive modeling process. Chapter 1 addresses data preprocessing, a necessary and vital step that has to take place before any modeling activity can be initiated. Data preprocessing greatly influences the degree to which pattern extraction is feasible and successful. The chapter strives to increase the awareness that data preprocessing is an important part of predictive analytics and a potential leverage to increase performance. Core preprocessing tasks and techniques are reviewed and some guidelines are provided on how to choose among alternative procedures. Furthermore, an empirical case study is undertaken to explore the relationship between prediction method, preprocessing, and forecasting accuracy. The results confirm a significant accuracy impact for certain preprocessing techniques and evidence that their effectiveness differs across prediction methods.

Chapter 2 zooms in upon text mining, a technique for assembling data when the data are in text format. The increasing amount of textual customer information that is stored in customer data warehouses leads to increased challenges and opportunities for marketing managers to better grasp the underlying customer behavior. However, marketing analytics often neglect this valuable type of information as it requires additional knowledge and effort to convert the text into a numeric representation suitable for subsequent processing. This chapter discusses the text mining process, and zooms into:

1. the text preprocessing phase that convert the textual consumer data into a high dimensional term-by-document matrix;
2. the dimension reduction techniques singular value decomposition and non-negative matrix factorization that group together related terms and projects them into a semantic space of lower dimensionality that could be used in traditional marketing analysis; and
3. the text mining applications published in top-tier marketing journals.

Chapter 3 shifts the focus to the actual model building process. Bayesian networks are popular within the fields of artificial intelligence and data mining due to their ability to support probabilistic reasoning from data with uncertainty. They can represent the co-related relationships among random variables and the conditional probabilities of each variable from a given data set. With a network structure at hand, people can conduct probabilistic inference to predict the outcome of some variables based on the values of other observed ones. The objective of the direct marketing modeling problem is to predict and rank potential buyers from the buying records of previous customers. The

customer list will be ranked according to each customer's likelihood of purchase. Bayesian networks can estimate the probability of a customer belonging to certain class(es) and are therefore suitable for many database marketing applications. For example, by assuming the estimated probability to be equal to the likelihood of purchase or response, they are suitable to handle the direct marketing problem. However, the databases containing the buying records of customers may contain missing values. This chapter gives an introduction to Bayesian networks and proposes a system for discovering Bayesian networks from incomplete databases in the presence of missing values. The authors apply it to a real-world direct marketing modeling problem, and compare the performance of the discovered Bayesian networks with other models obtained by other methods. In the comparison, the Bayesian networks learned by the proposed system outperform other models.

Chapter 4 discusses quantile regression and its relevance for database marketing. The simple yet well-performing and easily interpretable statistical methods such as linear and logistic regression account as gold standard methods in predictive modeling for marketing in both academic literature and business usage. In regression, an equation is sought describing the relationship between a number of independent variables and the mean of the dependent variable conditional upon the independent variables' values. Unfortunately, a mean is a strong simplification of reality as it is unable to unveil other characteristics of the underlying data distribution and might lead to incomplete or flawed conclusions when the conditional distribution is for instance highly skewed or contains outliers. Quantile regression is a generic approach that extends the mean regression model to a model specifying the relationship between covariates and any conditional quantile of the response variable of interest.

This chapter introduces the topic of quantile regression. A distinction is made between the frequentist and the Bayesian approach to estimate such models. Further, special attention is given to a recent development within this family of methods: binary quantile regression. Then, an elaborate section discusses the potential usage and advantages of quantile regression for database marketing through two case studies on customer lifetime value and customer churn prediction.

Chapter 5 sheds light upon the advantages of letting predictive models in database marketing join forces, whereby several models are combined into new, more flexible, and more powerful models. These so-called ensemble learners or multiple classifier systems have consistently emerged as winning entries in data mining contests, such as the Teradata/Duke CRM competition, KDD Cup or the Netflix Prize since many years. However, despite their strength and intuitive nature, their applications in real-life business are still scarce. This chapter untangles the topic of ensemble learning by first explaining their common structure, shared by the numerous algorithms that have been proposed within this category of statistical learners over recent years. Three intuitive arguments are presented to explain the potential of these methods to predict more accurately. The chapter continues with an elaborate overview of a selection of the most prominent and relevant ensemble learning algorithms for classification. Subsequently, it provides an overview and discussion of the academic literature on real-life database marketing applications in which ensemble learning was deployed, while a practical example is used to illustrate several concepts throughout the chapter. Special attention is given to two more advanced topics: (1) diversity, a key ingredient of any successful ensemble learner, and how it can be measured and assured; and (2) model interpretability.

The final chapter of the first part of this book (Chapter 6) shifts the focus to implementation by focusing on interpreting and operationalizing predictive models. Various different data mining techniques for marketing purposes are recently discussed in literature and have proven their excellence performance in a day-to-day business setting. Besides the search for optimal prediction performance, classification models should be intuitively correct and in accordance with the experts' knowledge. This chapter focuses on rule-based methods, that is, techniques which supplement the superior performance of black-box models with a set of insightful and comprehensible rules. These techniques open up these black-box models. The chapter summarizes the state-of-the-art rule-based models, describes the use of decision tables to visualize the extracted rules and concludes with an application in a churn prediction setting.

2.2 PART II: APPLICATIONS

Part II describes new applications in which the principles of database marketing can be applied in companies. Chapter 7 presents an integrated, comprehensive discussion of recommender systems. Recommender systems are software systems and statistical procedures used by firms to suggest ("recommend") products to their customers. A recommender system consists of data, a user model, and a selection model. A recommender system utilizes customer and product data to predict what product the customer is likely to prefer or purchase, and uses this prediction to select the product to be recommended to the customer. The data used to compute these predictions can pertain to the user, the product, or the user/product dyad. The chapter builds a general structure that shows how all three types of predictor data can be integrated into a "hybrid latent factor model." The model incorporates observed as well as unobserved user, product, and dyad data. It combines content-type user models that rely on observed predictors and collaborative filtering models that rely just on observed preferences or purchases. The chapter shows how the general model can be extended to binary preferences, missing data, unary data, buying context, and preference evolution. It then includes a discussion of estimation and selection models and closes with an overview of future research topics.

Chapter 8 examines strategic goals, tactics, and research issues related to marketing via mobile devices. The strategic goals examined include advertising and promotion, targeting, branding, and sales. The tactics that help achieve these goals include mobile web and applications ("apps"), mobile social media and social networks, location-based services, and mobile commerce. Mobile devices enhance the potential of companies to market to customers in real time more so than ever before. The chapter discusses these marketing tactics and reviews academic research relevant to them. This research provides key insights on how consumers use mobile devices, how consumers generate and consume user-generated content, how consumers select and use apps, how consumers use mobile devices as social media, the relevance of geography, and complementarity/substitution between mobile and non-mobile channels. While there is much more that needs to be learned, and the mobile platform is evolving, research to-date reveals the tremendous potential of mobile devices to enhance marketing effectiveness.

In Chapter 9, online display advertising and its targeting strategies are addressed. Online display advertising, or banner advertising, while being one of the earliest forms of advertising on the Internet, is still highly relevant today as expenditures have been consistently on the rise. Simultaneously however, this form of online advertising has

been troubled by steadily decreasing effectiveness in terms of click-through rates, a phenomenon often described as banner blindness. However, over the years, as strategies have been developed to gather increasing amounts of data and establish better metrics and analytics, online marketers have developed and refined their ability to target display ads to the most promising prospects. This chapter first provides an overview of metrics that are commonly used to evaluate the effectiveness of online display ads. Thereby, initially, a distinction is made between metrics that measure short-term and long-term effects. Further, the chapter discusses models that take into account immediate and long-term response simultaneously, and models that formally incorporate the multi-channel effect of online advertising. Finally, targeting strategies are discussed that deploy individual information to match ads and their most likely responders. Subsequently, targeting based upon user characteristics, geographical targeting, contextual targeting, and behavioral targeting are discussed.

Chapter 10 discusses paid search advertising, where Internet advertisers reach customers in the midst of their product search process. The chapter addresses the direct and indirect impacts of paid search. The direct effect is the immediate impact of a paid search ad, whereas indirect effects are longer term. The chapter reviews the history of paid search advertising and institutional issues such as the bidding process for ad placement. It then turns to a summary of empirical studies and models pertaining to direct effects, including the determinants of click-through rates and conversion. The chapter next discusses indirect effects including the impact of generic search on future branded search, the impact of click-through visits on future visits, the value of search advertising as a customer acquisition channel, and search ad copy design. The chapter concludes with a discussion of emerging topics such as the long tail in paid search, and the relationship between organic search and paid search click-throughs.

Chapter 11 summarizes how the firm can manage online social interactions. First, it describes the why and what of social interactions. That is, the chapter discusses what motivates consumers to share product recommendations and what product characteristics result in more word of mouth. The chapter then discusses issues related to social media metrics and data collection. It next proceeds to summarize existing research on the three roles that the firm can play in the management of social interactions:

1. observer;
2. influencer; and
3. participant.

Existing research suggests that the firm can measure the impact of social media interactions and successfully influence these interactions. While there is growing literature related to the role of the firm as observer and as influencer, the role of the firm as participant has not been studied extensively. Hence, the opportunities for impactful new research are greatest in this area.

Dynamic customer optimization models, discussed in Chapter 12, combine customer response models and optimization to determine what types of marketing to target to which customers at what time in order to maximize customer lifetime value. The fundamental premise of dynamic customer optimization is that marketing activities targeted in the current period should take into account the implications of these actions for marketing in subsequent periods. This chapter reviews the dynamic elements of

customer response models that come to play in dynamic customer optimization, and then discusses the fundamentals of dynamic optimization techniques. The chapter next reviews the evolution of the customer optimization field, starting with its roots in sales force management, proceeding to modern applications in catalogs, to more recent applications involving marketing tactics such as emails, sampling, and coupons, and free shipping. The last section of the chapter reviews several research papers, discussing the customer response model, the optimization, and the application. We conclude with a discussion of the promise and challenges of dynamic customer optimization models.

Chapter 13 focuses on how direct marketing practices in the non-profit sector differ from traditional direct marketing activities in the public and private sector. Using real-life case studies throughout the chapter, the author illustrates several direct marketing activities along the customer lifecycle, that is, donor acquisition, retention, and reactivation. Understanding charitable giving in its stage of the customer lifecycle is of crucial importance to optimize the donor database. More specifically, three questions are crucial: "To whom should the non-profit organizations send a donor invitation?"; "How to optimize or personalize the content of the mailing campaign?" and "Via which channels should the direct marketer target its donors?" This chapter concludes with an explanation of numerous campaign evaluation keywords.

References

Berry, L.L. (1983). *Relationship Marketing*. American Marketing Association, Chicago.

Blattberg, R.C., Kim B.-D. and Neslin, S.A. (2008). *Database Marketing: Analyzing and Managing Customers*. Springer, New York.

This page has been left blank intentionally

PART

Methods

This page has been left blank intentionally

CHAPTER 1

Data Preprocessing in Database Marketing: Tasks, Techniques, and Why They Matter

STEFAN LESSMANN

1 Introduction

The marketing domain has a long tradition of employing quantitative models to elicit and support decision making. For example, the well-known RFM (recency, frequency, monetary value) model has been used for decades by mail-order companies and charity organizations for targeting purposes (for example, Bauer, 1988). Nowadays, companies have access to vast amounts of customer-centric data. The data is gathered internally through a holistic use of information systems to support core business operations, and externally through partnering and exchanging data with suppliers, logistic service providers, and so on to improve efficiency and effectiveness of the overall supply chain. Finally, rich sets of data are collected from direct interactions with customers through various channels. The availability of such data has led to data-driven decision aids becoming ever more prominent in marketing.

Forecasting plays an important role in database marketing. Predictive models help marketers to anticipate future customer behavior and are routinely employed in a variety of applications including estimating customer lifetime value (for example, Bemmaor and Glady, 2011; Kumar, et al., 2006; Kumar, et al., 2008), fighting attrition through proactively identifying likely churners (for example, Glady, et al., 2009; Neslin, et al., 2006; Verbeke, et al., 2012), predicting customers' share of wallet (for example, Du, et al., 2007; Reichheld, 1996; Rosset, et al., 2007), or modeling customers' likelihood of responding to direct mail (for example, Bose and Xi, 2009; G. Cui, et al., 2006; O'Brien, 1994).

The vast majority of marketing prediction models ground on the principles of supervised learning. To estimate customers' likelihood of reacting to direct mail, for example, a functional relationship (that is, a model) between a set of independent variables, usually associated with the recency, frequency and monetary value of customers' past purchases as well as demographic customer information, and a zero-one dependent variable indicating whether or not a customer has responded to a past campaign, is

inferred (for example, Banslaben, 1992; Deichmann, et al., 2002; Malthouse, 2001). [1] In supervised learning, models are estimated (*learnt*) from past observations where attribute values *and* actual values of the target variable have been observed (for example, Hastie, et al., 2009). The resulting model facilitates predicting the target variable when only the attribute values are observable; for example, estimating how likely it is that customers (with known attribute values) respond to a future mailing if they are solicited.

Much research has concentrated on examining the effectiveness of alternative prediction methods such as logit models, neural networks, or tree-based ensemble learners for different marketing tasks (for example, Coussement, et al., 2010; Coussement and Van den Poel, 2008; Cui and Curry, 2005; Curry and Moutinho, 1993; De Bock and Van den Poel, 2011; Ghosh, et al., 1984; Kim, et al., 2005; Lemmens and Croux, 2006; Neslin, et al., 2006; West, et al., 1997). From a practitioner's point of view, devising a prediction model might not be the biggest challenge since many powerful techniques are readily available in standard software packages. Instead, the task of collecting and integrating the data needed for any modeling activity is often the most time-consuming and costly step within industrial data mining projects (for example, Berry and Linoff, 2011). In this sense, data collection/integration and model building[2]/development can be regarded as particularly well explored and understood from a practical and academic perspective, respectively.

This chapter deals with data preprocessing, a modeling phase in between data collection/integration and model building, which has received much less attention in both academia and industry. The objective of data preprocessing is to convert raw data into a format that facilitates the application of quantitative prediction methods *and* aids the model building algorithm in extracting predictive information from the independent variables. In particular, the choices made within data preprocessing have a significant influence upon the accuracy of prediction models (Crone, et al., 2006; Neslin, et al., 2006).

The chapter gives an overview of data preprocessing for database marketing and illustrates the choices that need to be made in this modeling stage. This is to increase awareness for data preprocessing as an integral part of predictive modeling and a potential opportunity to increase forecasting accuracy. In addition, the exposition shall equip readers with a sound understanding of alternative preprocessing techniques and their respective merits and demerits. Finally, a case study is undertaken to:

1. demonstrate empirically the impact of preprocessing on predictive accuracy;
2. explore the intricate relationship between prediction and preprocessing methods in some detail; and
3. provide some guidance on when a specific form of preprocessing is effective.

The remainder of the chapter is organized as follows: Section 2 briefly reviews the overall data mining process, to give a context for data preprocessing activities and techniques. These are elaborated in Section 3. The results of the empirical case study are presented in Section 4, before conclusions are drawn in Section 5.

---

1 The dependent and independent variables are known by many names in the literature. In this chapter, the dependent variable is referred to as the target variable. Independent variables are synonymously referred to as attributes or features.

2 Also termed model estimation in the statistics literature.