Chapter 1

Introduction

Various types of data have been stored in a lot of fields, such as science, engineering, economics, agriculture, finance, etc., and so we are in the era of *big data*. The scale advantage of abundant data should be used effectively, to make our life more comfortable and richer, for example, more advanced sciences and technologies, more affordable healthcare, higher efficiency of business activities, and so on. One example is marketing in business, where a variety of data on products, customers, services, companies and their interactions, like user-item purchase records and social relationships between customers, have been already obtained and further increasing [23]. Numerous aspects of marketing can be application targets of so-called *data-driven* approaches, such as business analytics, data science, data mining and machine learning, which lead to predictive *marketing analytics* [111, 16], or more generally artificial intelligence [46, 35, 97]. Typically, recommendation has already long history, being started with rather a simple idea of, for a target customer, finding similar customers with previous purchase records and suggesting the target customer items which are not yet purchased but bought by most of similar customers. This intuitive concept, called *collaborative filtering* [47], behind recommendation would be reasonable, consistent with the thoughts owned by marketers, for a long time. Customer churn analysis would be another example, which is more suitable to be addressed by machine learning to predict current membership customers to potentially cancel the contract in the near future, by using past and present customer data and records [2]. More concretely, customer churn prediction is a proper application of binary classification, providing machine learning research with an interesting, challenging problem of class imbalanceness on training data due to a limited number of past (but not current) customers. These two application examples are just a part of iceberg above water, and other applications have been addressed and at the same time, numerous opportunities would be still waiting for data-driven approaches to be applied in marketing. In other words, the current applications including the above two examples might be rather exceptional, since they have been already well considered and developed by data-driven techniques, particularly machine learning, but we think that a plenty of promising situations to be solved by machine learning still remain unsolved, untouched or even unconceived in marketing.

However recent issues in business magazines and academic journals on marketing and consumer research show that machine learning-based work for new applications are extensively emerging in marketing, particularly using the data, especially those called *user-generated content* (UGC) from social networks or social media as the main input. A short list of examples of the new applications using images and text of social networks (or social media) includes retrieving online social ties [114], extracting brand information [51, 66, 55, 110, 49], understanding consumer needs [101], consumer sharing [108] and consumer behaviors on advertising [68], detecting key terms relevant to sentiment patterns [109], measuring social media influencer index [6], to name just a few. Also online pricing is definitely an interesting issue in marketing [93, 74]. In fact these work mostly have appeared in the last few years, and before that it was rare to meet such papers. We can say that promising applications now appear here and there in marketing, making use of the predictive power of machine learning. A noteworthy feature of these work is that the used data are obtained through modern information sources, i.e. internet, such as social media or social networks, and the applications are motivated by those data. Although however first they look totally new as applications, we can then notice that these applications can also use the traditional data, such as survey data by customers over products which eventually has the same format and information as the data obtained from reviews in E-commerce sites or blogs. In other words, currently marketing researchers apply machine learning to the modern data directly, skipping doing so over the traditional data, though the same applications can be made from the data in traditional marketing, such as target marketing and relationship marketing. This indicates that it would be important to consider to use machine learning over the traditional framework of marketing, particularly over the data already obtained and accumulated. Another characteristic point of note on machine learning in marketing currently is that text mining is well used to capture knowledge (useful for marketing) from particularly online text data, so-called *electric word-of-mouth* (eWOM), including blogs and reviews. In fact, a lot of work of machine learning in marketing, including those shown in the above, are in this line of research [49, 34, 78, 96, 24]. Text mining in marketing however would be rather more natural language processing than machine learning, and so will not be touched in this book.

Machine learning can be used for not only analyzing the stored data to understand the past results but also for predicting the future of the changing world of marketing [57]. Also as mentioned in the above briefly, machine learning can be applied to not only recent online data but also traditionally obtained data in marketing. We can then say that now the two sides, i.e. machine learners and marketers, may understand the other side more to explore and solve various new problems/applications in marketing by machine learning. A key point is to build a proper machine learning approach for each application. This stage would need a proper guide, which presents ways of defining machine learning problems and developing methods for solving the problems from data given in marketing, particularly not necessarily only new data like social networks and related data but also traditional data like customer demographic data, user survey, purchase records and user-item matrix for recommendation. In fact machine learners can design, for a given problem, a model and its algorithm flexibly, according to given data or the assumption behind the problem (even under the same problem setting), instead of just setting up a problem to which existing methods can be applied. This manner of understanding machine learning would be key for those who, in the application side, are likely to apply existing machine learning approaches straightforwardly to each problem setting. In fact again as mentioned above, although text mining has been well used in marketing for understanding eWOM, text mining would be also more likely to use existing machine learning methods as tools rather directly, e.g. text classification in marketing [40], tagging online contents automatically [85], and also text analysis for consumer research [48]. Thus again text mining will not be mentioned in this book.

There already exist books on the applications of machine learning, data mining or rather data science (business analytics) to marketing [28, 64]. Their contribution to marketers, particularly practitioners working on real data, would be tremendous. For example, [28] provides a comprehensive, practical, step-by-step guide of applying data science techniques to market segmentation, showing even codes in R. The authors of these books, basically from the application side, assume that machine learning methods are already given, and then each problem is set so that machine learning algorithm can be easily applied. This would be reasonable. On the other hand, however, an important point would be, again, to show how machine learning methods can be (uniquely) designed for a given application, and can be changed, according to available data or the assumption behind the application. Over all we can again emphasize that machine learning models can be designed by the idea/ assumption behind the problem setting.

The objective of this book is primarily to show the most appropriate (and simplest) machine learning method for the given problem and its assumption, and then present a way of exploring proper changes of machine learning methods, according to the variation of data, problem settings and/or assumptions behind even a single problem setting. The machine learning methods in this book are kept as simple as possible, particularly comparing with the recent, more complicated, black-box-type methods, such as deep neural networks. We think that by using such simple models, readers will understand each model and its variations more, particularly reasonably the model being modified, according to given data and the assumption behind the problem. We hope that this book will be helpful for marketers to understand machine learning models in the manner described (Interested readers can refer to machine learning books more, e.g. [70]), and for machine learners to understand the idea and points of applications in marketing.

In this book, we focus on two mainstream (and traditional) concepts of marketing: *target marketing* and *relationship marketing*, which have been established already and now well matured in marketing. In fact both are key ideas generated and developed in the modern marketing history. Additionally we consider database marketing, which would be applications more suitable for machine learning. We explore the possibility of developing a machine learning method for each step (or aspect) of these two concepts, assuming that we have a plenty of related data in marketing, such as those for customers, products, manufactures, their interactions, such as purchase records, and whatever.

The organization of this book can be given below.

Chapters 1 to 4 are the introductory part of this book, providing basic knowledge and information, which must be well understood for learning both machine learning and marketing.

Chapter 1, this chapter, purely introduces the ideas behind this book. Chapter 2 provides the concepts and terminologies in machine learning first and then marketing. Again this book focuses on two main concepts of marketing, i.e. target marketing and relationship marketing. Chapter 3 introduces the fundamentals of target marketing, which consists of three steps, called the STP (segmentation, targeting and positioning) strategy. These three steps are described in detail but concisely. Also the STP strategy in general uses several interesting ideas, such as SWOT (strengths, weaknesses, opportunities and threats) analysis and perceptual map. This chapter describes these ideas briefly but clearly by using several examples for each of them. Chapter 4 describes the basics of relationship marketing. In particular, this chapter focuses on several topics, such as RFM (recency, frequency and monetary) analysis, customer loyalty satisfaction, retention marketing (marketing funnel), which are described with some examples.

Chapters 5 to 8 are the main part of this book, presenting various machine learning approaches for marketing applications and their variants. The key point is we can change, for example, the data, problem setting and/or assumption behind the problem, and for each time, we can develop a machine learning method for each particular problem setting. We then describe such diverse models in these chapters, resulting in showing diverse machine learning approaches in these chapters.

Chapter 5 starts with fundamental machine learning concepts and approaches, which are required to address the problems in marketing to be introduced in later chapters. The machine learning paradigms in this chapter can be first classified into three concepts: regular machine learning, feature learning and kernel learning. In particular, regular machine learning focuses on two data types: vectors and nodes in a graph, and then for each type, three machine learning concepts, i.e. supervised, unsupervised and semisupervised learning, and so methods for totally six combinations (two data types $\times$ three learning concepts) are described. Feature learning has two types, feature selection and feature generation (dimensionality reduction), which both are described concisely. Finally for kernel learning, a standard procedure to make regular machine learning procedure kernelized into the corresponding method in kernel learning.

In Chapter 6, we focus on target marketing or the STP (segmentation, targeting and positioning) strategy, and explore the possibility of developing machine learning methods or applying the ideas of machine learning, for many aspects of the STP strategy. For example, clustering has been applied to segmentation, while in this book, we consider a variety of machine learning approaches for segmentation, depending on the data and assumptions we can consider. Also we will think about machine learning-based segment evaluation and SWOT analysis. Furthermore we present machine learning methods for generating perceptual maps, depending on the assumptions behind the maps, which are used for positioning the products, brands and companies. Finally we consider an interesting problem setting of finding the area in a perceptual map (or any dimensional space) which have no competitive products but a lot of customers. We then present a simple machine learning model and algorithm to address this problem.

Chapter 7 focuses on relationship marketing, and explores the possibility of building ML models to apply each of all possible parts of relationship marketing. This chapter contributes to the three aspects of relationship marketing: customer relationship management (CRM), retention marketing and market communications. In CRM, we consider three problems: learning user-item (customer-product) matrix, detecting profitable customers (automatic RFM (recency, frequency, monetary) analysis) and customer churn analysis. Then for retention marketing, we set up a general setting of problems, considering flexible transitions between different stages in marketing funnel, instead of incremental or progressive promotion through the marketing funnel. For this setting, we present two types of ML solutions. Finally in this chapter, we address a problem of optimally selecting communication channels among a wide variety of choices on current market channels, assuming that a good amount of data can be given to optimize the built model.

Finally Chapter 8 focuses on settings, in which not only one but multiple matrices are given as input. For example, given two or more matrices which always share one dimension, e.g. rows for instances, and in which the other dimension can be changed, like that different feature sets can be given, we may then want analyze these matrices, for finding interactions between different feature sets of different matrices or capturing common factors shared by different feature sets, etc. In this chapter, under the setting of given multiple matrices, we consider four types of ML settings: 1) supervised, 2) unsupervised (clustering), 3) unsupervised (factorization) and 4) learning interactions between features. Also we change the setting of given multiple matrices under the above four machine learning settings, while our focus goes more on to matrix factorization eventually, so-called *collaborative matrix factorization* for given multiple matrices [115]. Overall in this chapter, we present numerous types of collaborative matrix factorization, considering general combinations with an arbitrary number of matrices, as a representative data integrative approach.