Birkhäuser

Applied and Numerical Harmonic Analysis

For further volumes:
http://www.springer.com/series/4968

Alexander Paprotny • Michael Thess

Realtime Data Mining

Self-Learning Techniques for Recommendation Engines

Birkhäuser

Alexander Paprotny
Research and Development
prudsys AG
Berlin, Germany

Michael Thess
Research and Development
prudsys AG
Chemnitz, Germany

Corrected at 2nd printing 2014

ISSN 2296-5009 ISSN 2296-5017 (electronic)
ISBN 978-3-319-01320-6 ISBN 978-3-319-01321-3 (eBook)
DOI 10.1007/978-3-319-01321-3
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013953342

Mathematics Subject Classification (2010): 68T05, 68Q32, 90C40, 65C60, 62-07

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.birkhauser-science.com)

ANHA Series Preface

The Applied and Numerical Harmonic Analysis (ANHA) book series aims to provide the engineering, mathematical, and scientific communities with significant developments in harmonic analysis, ranging from abstract harmonic analysis to basic applications. The title of the series reflects the importance of applications and numerical implementation, but richness and relevance of applications and implementation depend fundamentally on the structure and depth of theoretical underpinnings. Thus, from our point of view, the interleaving of theory and applications and their creative symbiotic evolution is axiomatic.

Harmonic analysis is a wellspring of ideas and applicability that has flourished, developed, and deepened over time within many disciplines and by means of creative cross-fertilization with diverse areas. The intricate and fundamental relationship between harmonic analysis and fields such as signal processing, partial differential equations (PDEs), and image processing is reflected in our state-of-the-art *ANHA* series.

Our vision of modern harmonic analysis includes mathematical areas such as wavelet theory, Banach algebras, classical Fourier analysis, time-frequency analysis, and fractal geometry as well as the diverse topics that impinge on them.

For example, wavelet theory can be considered an appropriate tool to deal with some basic problems in digital signal processing, speech and image processing, geophysics, pattern recognition, biomedical engineering, and turbulence. These areas implement the latest technology from sampling methods on surfaces to fast algorithms and computer vision methods. The underlying mathematics of wavelet theory depends not only on classical Fourier analysis but also on ideas from abstract harmonic analysis, including von Neumann algebras and the affine group. This leads to a study of the Heisenberg group and its relationship to Gabor systems, and of the metaplectic group for a meaningful interaction of signal decomposition methods. The unifying influence of wavelet theory in the aforementioned topics illustrates the justification for providing a means for centralizing and disseminating information from the broader, but still focused, area of harmonic analysis. This will be a key role of *ANHA*. We intend to publish with the scope and interaction that such a host of issues demand.

Along with our commitment to publish mathematically significant works at the frontiers of harmonic analysis, we have a comparably strong commitment to publish major advances in the following applicable topics in which harmonic analysis plays a substantial role:

Antenna theory	*Prediction theory*
Biomedical signal processing	*Radar applications*
Digital signal processing	*Sampling theory*
Fast algorithms	*Spectral estimation*
Gabor theory and applications	*Speech processing*
Image processing	*Time-frequency and time-scale analysis*
Numerical partial differential equations	*Wavelet theory*

The above point of view for the *ANHA* book series is inspired by the history of Fourier analysis itself, whose tentacles reach into so many fields.

In the last two centuries, Fourier analysis has had a major impact on the development of mathematics, on the understanding of many engineering and scientific phenomena, and on the solution of some of the most important problems in mathematics and the sciences. Historically, Fourier series were developed in the analysis of some of the classical PDEs of mathematical physics; these series were used to solve such equations. In order to understand Fourier series and the kinds of solutions they could represent, some of the most basic notions of analysis were defined, for example, the concept of "function." Since the coefficients of Fourier series are integrals, it is no surprise that Riemann integrals were conceived to deal with uniqueness properties of trigonometric series. Cantor's set theory was also developed because of such uniqueness questions.

A basic problem in Fourier analysis is to show how complicated phenomena, such as sound waves, can be described in terms of elementary harmonics. There are two aspects of this problem: first, to find, or even define properly, the harmonics or spectrum of a given phenomenon, for example, the spectroscopy problem in optics; second, to determine which phenomena can be constructed from given classes of harmonics, as done, for example, by the mechanical synthesizers in tidal analysis.

Fourier analysis is also the natural setting for many other problems in engineering, mathematics, and the sciences. For example, Wiener's Tauberian theorem in Fourier analysis not only characterizes the behavior of the prime numbers but also provides the proper notion of spectrum for phenomena such as white light; this latter process leads to the Fourier analysis associated with correlation functions in filtering and prediction problems, and these problems, in turn, deal naturally with Hardy spaces in the theory of complex variables.

Nowadays, some of the theory of PDEs has given way to the study of Fourier integral operators. Problems in antenna theory are studied in terms of unimodular trigonometric polynomials. Applications of Fourier analysis abound in signal processing, whether with the fast Fourier transform (FFT) or filter design or the adaptive modeling inherent in time-frequency-scale methods such as wavelet theory. The coherent states of mathematical physics are translated and modulated Fourier transforms, and these are used, in conjunction with the uncertainty principle, for dealing with signal reconstruction in communications theory. We are back to the raison d'être of the *ANHA* series!

University of Maryland
College Park

John J. Benedetto
Series Editor

Reprinted with the permission of Frank Nathan

Preface

The area of realtime data mining is currently developing at an exceptionally dynamic pace. Realtime data mining systems are the counterpart of today's "classic" data mining systems. Whereas the latter learn from historical data and then use it to deduce necessary actions, realtime analytics systems learn and act continuously and autonomously. In the vanguard of these new analytics systems are *recommendation engines* (REs). They are principally found on the Internet, where all information is available in real time and an immediate feedback is guaranteed.

In this book, we describe novel mathematical concepts for recommendation engines based on realtime learning. These feature a sound mathematical framework which unifies approaches based on control and learning theories, tensor factorization, and hierarchical methods. Furthermore, they present promising results of numerous experiments on real-world data. Thus, the book introduces and demystifies this concept of "realtime thinking" for a specific application—recommendation engines. Additionally, the book provides useful knowledge about recommendation engines such as verification of results in A/B tests including calculation of confidence intervals, coding examples, and further research directions.

The main goal of the research presented in the book consists of devising a sound and effective mathematical and computational framework for automatic adaptive recommendation engines. Most importantly, we introduce an altogether novel control-theoretic approach to recommendation based on considering the customer of an (online) shop as a dynamic system upon which the recommendation engine acts as a closed-loop control system, the objective of which is maximizing the incurred reward (e.g., revenue). Besides that, we also cover classical data mining-based approaches and develop efficient numerical procedures for computing and, especially, updating the underlying matrix and tensor decompositions. Furthermore, we take a step toward a framework that unifies the two approaches, that is, the classical and the control-theoretic one. In summary, the book proposes a very modern approach to realtime analytics and includes a lot of new material.

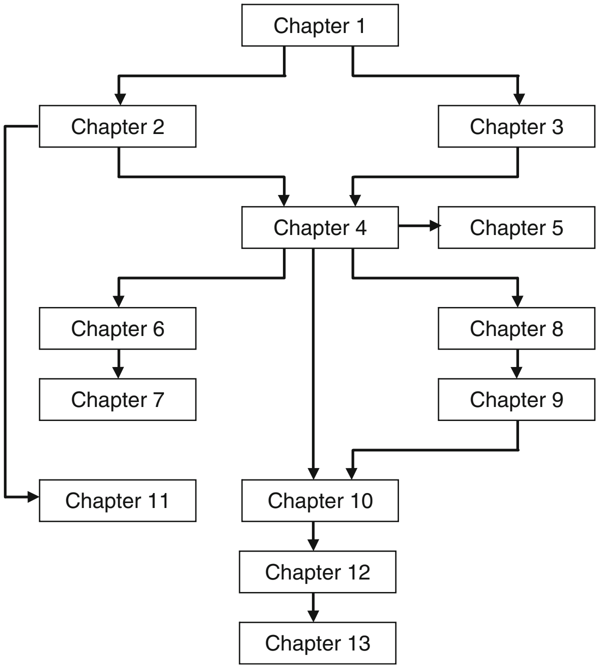
Currently, most books about recommendation engines focus on traditional techniques, such as collaborative filtering, basket analysis, and content-based recommendations. Recommendations are considered from a prediction point of view only, that is, the recommendation task is reduced to the prediction of content that the user is going to select with highest probability anyway. In contrast, in our book we consider recommendations as a control-theoretic problem by investigating the interaction of analysis and action. At this, an optimization problem with respect to maximum reward is considered.

Another important frequently recurring theme in our train of thought is that of hierarchical approaches. In recent decades, methods that capture and take into account effects at different scales have turned out to be a key ingredient to successfully tackling complex problems in signal processing and numerical solution of partial differential equations. Supported by the evidence that we shall present in this book, we strongly conjecture that this paradigm may give rise to major improvements in the efficiency of computational procedures deployed in the framework of realtime recommendation engines. We therefore would like to stress that this book is also a step toward introducing harmonic thinking in the theory and practice of recommendation engines.

The book targets, on one hand, computer scientists and specialists in machine learning, especially from the area of recommendation systems, because it conveys a new way of realtime thinking especially by considering recommendation tasks as control-theoretic problems. On the other hand, the book may be of considerable interest to application-oriented mathematicians, because it consistently combines some of the most promising mathematical areas, namely, control theory, multilevel approximation, and tensor factorization.

Owing to the complexity of the subject, the book cannot go into all the details of the mathematical theory, let alone its implementation. Nevertheless, it sets out the basic assumptions and tools that are needed for an understanding of the theory. In some areas of fundamental importance, we also offer more detailed mathematical examples. Overall, however, we have tried to keep the mathematical illustrations short and to the point.

The document structure is as follows. Chapter 1 offers a general introduction to methods of realtime analytics and sets out their advantages and disadvantages as compared with conventional analytics methods, which learn only from historical data. Chapter 2 describes conventional approaches for recommendation engines and shows how their inherently static methodology is their main weak point. The use of realtime analytics methods is suggested as a way of overcoming precisely this problem and, specifically, reinforcement learning (RL), one of the very newest disciplines, which models the interplay of analysis and action. Chapter 3 provides a brief introduction to RL, while Chap. 4 applies this knowledge to recommendation engines. There are still a number of fundamental problems to resolve, however, requiring the introduction of some additional empirical assumptions. This is done in Chap. 5, resulting in a complete RL-based approach for recommendation engines.



The next chapters are devoted to improve our solution, especially concerning stability and speed of convergence. Thus, in Chap. 6 we study hierarchical methods and add a hierarchical convergence accelerator to further boost the learning speed. Chapter 7 represents an extension to the topic of hierarchical methods where a powerful adaptive scoring technique is described—sparse grids. For a better exploitation of the data in the calculation of recommendations, in Chap. 8 we introduce matrix factorization techniques along with some adaptive implementation. Using the tensor concept, Chap. 9 extends the factorization to the high-dimensional case. This enables us to combine hierarchical RL with adaptive tensor factorization in order to include additional dimensions into the realtime calculation of recommendations. This "big picture," which is still in the very beginning, is described in Chap. 10 and concludes the technical description of our new recommendation approach.

In Chap. 11, we discuss statistically rigorous methods for measuring the success of recommendation engines. Chapter 12 is devoted to the prudsys XELOPES library which implements most of the algorithms described in this book and provides a powerful infrastructure for realtime learning. Finally, in Chap. 13 we summarize the main elements covered in the book.

Parts of the book provide an easily understandable introduction to realtime recommendations and do not require deep mathematical knowledge. Especially, this applies to Chaps. 1 and 2 as well as Chaps. 11, 12, and 13. Chapters 3, 4, and 5 are devoted to reinforcement learning and assume basic knowledge of algebra and statistics. In contrast, Chaps. 6, 7, 8, 9, and 10 address mathematically more experienced readers and require solid knowledge of linear algebra and analysis.

Acknowledgements

We would like to thank André Müller and Sven Gehre for their assistance in the tensor factorization chapter and Jochen Garcke for his critical review of the manuscript. We would also like to thank Holm Sieber for his help in the mathematical treatment of multiple recommendations and, additionally, Toni Volkmer for deriving the confidence intervals of revenue increase. Further we would like to thank the many reviewers who provided us with critical comments and suggestions. In particular, we would like to mention Jens Scholz, Tina Stopp, Gerard Zenker, and Brian Craig, as well as the anonymous reviewers allocated by the publisher.

Berlin, Germany Alexander Paprotny

Contents

Summary of Notation

Number Sets

$\mathbf{N}$	The set of natural numbers
$\mathbf{N}_0$	The set of natural numbers with 0
$\mathfrak{R}$	The set of real numbers

Set Operators and Relations

$\cup$	Union
$\cap$	Intersection
$\backslash$	Exclusion
$\prod_{i \in \mathfrak{I}} X_i$	Cartesian product of the sets X_i
$\subset$	Subset or equal

Discrete Sets and Graphs

$\underline{n}$	The set $\{1, \ldots, n\}$
$\lvert S \rvert$	Cardinality of the set S
G	A partition of an index set
$\Gamma = (V, E)$	Directed graph with vertex set V and edge set E
Γ_G	Aggregation of the graph Γ w.r.t. the partition G
$\Gamma\|_S$	Restriction of the graph $\Gamma = (V, E)$ to the set $S \subset V$

Spaces, Subsets of Spaces

$\mathfrak{R}^n$	Vector space of dimension n over $\mathfrak{R}$
$\mathfrak{R}^{n \times m}$	Matrix space of dimension $n \times m$ over $\mathfrak{R}$

$\mathfrak{R}^n_{\geq 0}, \mathfrak{R}^n_{>0}$	The sets of nonnegative and positive vectors of length n, respectively
$\mathfrak{R}^n_{>O}\left(\mathfrak{R}^n_{\geq O}\right)$	The sets of positive (nonnegative) $n \times n$ matrices
$\mathfrak{R}^n_{\succ O}\left(\mathfrak{R}^n_{\succeq O}\right)$	The sets of symmetric positive (semi-) definite $n \times n$ matrices
$\mathfrak{R}^S$	Space of functions $S \to \mathfrak{R}$, isomorphic to $\mathfrak{R}^{\lvert S\rvert}$
$V^\perp <\cdot,\cdot>$	Orthogonal space of V w.r.t. the inner product $<\cdot,\cdot>$

Subspaces Related to Matrices

ran A	Range space of A
ker A	Kernel (null space) of A

Components of Matrices and Vectors

v_i	i-th component of the vector v
a_{ij}, $(A)_{ij}$	Component in ith row and jth column of matrix A
$a^{(j)}$	j-th column of A

Operations on Matrices

A^T	Transpose of A
$\oplus$	Direct sum
$\otimes$	Kronecker product
$\sigma(A)$	Spectrum of A
$\rho(A)$	Spectral radius of A
$sub(A)$	Subdominant radius of A
$rank\ A$	Dimension of the range space of A

Inner Products and Norms

$<\cdot,\cdot>$	Generic inner product, also canonical inner product x^Ty
$<\cdot,\cdot>_S$	Inner product induced by the matrix S, $<x, y>_S = x^TSy$
$\lVert\cdot\rVert$	Generic norm
$\lVert\cdot\rVert^*$	Operator norm induced by $\lVert\cdot\rVert$
$\lVert\cdot\rVert_1$	l_1-norm
$\lVert\cdot\rVert_2$	l_2-norm
$\lVert\cdot\rVert_\infty$	l_∞-norm (max-norm)
$\lVert\cdot\rVert_w$	w-weighted max-norm
$\lVert\cdot\rVert_*$	Nuclear norm

Matrix Inverses

A^{-1} Algebraic inverse of A
A^{+} Moore-Penrose inverse of A
A^{+S} Moore-Penrose inverse of A w.r.t. $<\cdot,\cdot>_S$
A^{+w} Moore-Penrose inverse of A w.r.t. $<\cdot,\cdot>_{diag(w)}$

Important Vectors and Matrices

I_n $n \times n$ identity matrix
I Identity matrix (follows from context)
$O_{m,n}$ $m \times n$ matrix of all zeros
O_n $n \times n$ matrix of all zeros
O Matrix of all zeros (dimension follows from context)
$\mathbf{1}_n$ Vector of length n of all ones
$\mathbf{1}$ Vector of all ones (dimension follows from context)
$e_n^{(i)}$ The vector $\left(e_n{}^{(i)}\right)_j = \sigma_{ij},\ i,\ j \in \underline{n}$
$e^{(i)}$ The vector $e_n^{(i)}$ (n follows from context)
$\prod$ Projector
b Right-hand side of a system of linear equations
A Coefficient matrix of a system of linear equations
x^* Solution of a system of linear equations

Dynamic Programming

M Markov decision process (MDP)
π Policy
$\prod_M$ Set of all policies for the MDP M
M_π Markov chain induced by policy $\pi \in \prod_M$
S State space
A Action space
$A(s)$ Action set in state s
p_{ij}^a Probability of state transition from i to j given action a
r_{ij}^a Reward of state transition from i to j given action a
r_{ij} Reward of state transition from i to j (action follows from context)
P Transition probability tensor
R Transition reward tensor
P^π Transition probability matrix of Markov chain M_π
R^π Transition reward matrix of Markov chain M_π
r^π Transition reward of Markov chain M_π
v^π State-value function corresponding to the policy π

q^{π}	Action-value function corresponding to the policy π
$d_t(\cdot)$, $d_t^k(\cdot)$	Temporal-difference error at t (and iteration step k)
$z_t(\cdot)$, $z_t^k(\cdot)$	Eligibility trace at t (and iteration step k)
γ	Discount rate
ε	Probability of random action in ε-greedy policy
α, β	Step-size parameters
λ	Decay-rate parameter for eligibility traces

DP for Recommendations

s_a	State associated with recommendation a
a_s	Recommendation associated with state s
$S_{A(s)}$	State set associated with all m possible recommendations
$\overline{a}$	Composite recommendation of k recommendations: $\overline{a} = (a_1, \ldots a_k)$
$S_{\overline{a}}$	State set associated with all recommendations: $S_{\overline{a}} = \{s_1\} \cup \ldots \cup \{s_k\}$
$\prod_{\overline{a}}$	Transition probabilities of recommendation states: $\prod_{\overline{a}} = \{p_{ss'}\}_{s' \in S_{\overline{a}}}$

Multilevel

I_l^m	Level l-to-m interpolation/restriction operator
I_{l+1}^l	Interpolation (prolongation) operator from level $l+1$ to level l
I_l^{l+1}	Restriction operator from level l to level $l+1$
L	Interpolation (prolongation) operator I_1^0
R	Restriction operator I_0^1

Tensors

$\mathbf{i}$	Multi-index
$\underline{\mathbf{n}}$	Set of multi-indexes
$(\mathbf{m},\mathbf{n})$	Concatenation of multi-indexes
$\circ$	Outer vector product
$\otimes$	(Outer) tensor product
$\otimes_\delta$	Multilinear (concatenated) product over δ indexes
$\times_p$	Multilinear p-mode product with matrix (multilinear product with $\delta = 1$)
$\mathbf{i}^{(k)}$	Multi-index without k-th coordinate
$\underline{\mathbf{n}}^{(k)}$	Set of multi-indexes without k-th coordinate
$\overline{A}^{(k)}$	k-Mode matricization of A

Miscellaneous

δ_{ij}	Kronecker delta
diag(v)	Diagonal matrix with components of v on the diagonal
~, $\propto$	Proportional
$f\,\|_X$	Restriction of the function f to the set S
$\text{argmin}_{x \in X} f(x)$	The set of minimizers of the function $f\,\|_X$
$\text{argmax}_{x \in X} f(x)$	The set of maximizers of the function $f\,\|_X$
$V \perp_{<\cdot,\cdot>} W$	V is orthogonal to W in terms of the inner product $<\cdot,\cdot>$