Multi-Armed Bandits

Theory and Applications to Online Learning in Networks

Synthesis Lectures on Communication Networks

Synthesis Lectures on Communication Networks is an ongoing series of 75- to 150-page publications on topics on the design, implementation, and management of communication networks. Each lecture is a self-contained presentation of one topic by a leading expert. The topics range from algorithms to hardware implementations and cover a broad spectrum of issues from security to multiple-access protocols. The series addresses technologies from sensor networks to reconfigurable optical networks.
The series is designed to:

- Provide the best available presentations of important aspects of communication networks.
- Help engineers and advanced students keep up with recent developments in a rapidly evolving technology.
- Facilitate the development of courses in this field

Multi-Armed Bandits: Theory and Applications to Online Learning in Networks
Qing Zhao
2019

Diffusion Source Localization in Large Networks
Lei Ying and Kai Zhu
2018

Communications Networks: A Concise Introduction, Second Edition
Jean Walrand and Shyam Parekh
2017

BATS Codes: Theory and Practice
Shenghao Yang and Raymond W. Yeung
2017

Analytical Methods for Network Congestion Control
Steven H. Low
2017

Advances in Multi-Channel Resource Allocation: Throughput, Delay, and Complexity
Bo Ji, Xiaojun Lin, and Ness B. Shroff
2016

A Primer on Physical-Layer Network Coding
Soung Chang Liew, Lu Lu, and Shengli Zhang
2015

Sharing Network Resources
Abhay Parekh and Jean Walrand
2014

Wireless Network Pricing
Jianwei Huang and Lin Gao
2013

Performance Modeling, Stochastic Networks, and Statistical Multiplexing, Second Edition
Ravi R. Mazumdar
2013

Packets with Deadlines: A Framework for Real-Time Wireless Networks
I-Hong Hou and P.R. Kumar
2013

Energy-Efficient Scheduling under Delay Constraints for Wireless Networks
Randall Berry, Eytan Modiano, and Murtaza Zafer
2012

NS Simulator for Beginners
Eitan Altman and Tania Jiménez
2012

Network Games: Theory, Models, and Dynamics
Ishai Menache and Asuman Ozdaglar
2011

An Introduction to Models of Online Peer-to-Peer Social Networking
George Kesidis
2010

Stochastic Network Optimization with Application to Communication and Queueing Systems
Michael J. Neely
2010

Scheduling and Congestion Control for Wireless and Processing Networks
Libin Jiang and Jean Walrand
2010

Performance Modeling of Communication Networks with Markov Chains
Jeonghoon Mo
2010

Communication Networks: A Concise Introduction
Jean Walrand and Shyam Parekh
2010

Path Problems in Networks
John S. Baras and George Theodorakopoulos
2010

Performance Modeling, Loss Networks, and Statistical Multiplexing
Ravi R. Mazumdar
2009

Network Simulation
Richard M. Fujimoto, Kalyan S. Perumalla, and George F. Riley
2006

Multi-Armed Bandits: Theory and Applications to Online Learning in Networks

Qing Zhao

ISBN: 978-3-031-79288-5 paperback
ISBN: 978-3-031-79289-2 ebook
ISBN: 978-3-031-79290-8 hardcover

DOI 10.1007/978-3-031-79289-2

A Publication in the Springer series
SYNTHESIS LECTURES ON COMMUNICATION NETWORKS

Lecture #22
Series Editor: R. Srikant, *University of Illinois at Urbana-Champaign*
Founding Editor Emeritus: Jean Walrand, *University of California, Berkeley*
Series ISSN
Print 1935-4185 Electronic 1935-4193

Multi-Armed Bandits

Theory and Applications to Online Learning in Networks

Qing Zhao
Cornell University

SYNTHESIS LECTURES ON COMMUNICATION NETWORKS #22

ABSTRACT

Multi-armed bandit problems pertain to optimal sequential decision making and learning in unknown environments. Since the first bandit problem posed by Thompson in 1933 for the application of clinical trials, bandit problems have enjoyed lasting attention from multiple research communities and have found a wide range of applications across diverse domains. This book covers classic results and recent development on both Bayesian and frequentist bandit problems. We start in Chapter 1 with a brief overview on the history of bandit problems, contrasting the two schools—Bayesian and frequentist—of approaches and highlighting foundational results and key applications. Chapters 2 and 4 cover, respectively, the canonical Bayesian and frequentist bandit models. In Chapters 3 and 5, we discuss major variants of the canonical bandit models that lead to new directions, bring in new techniques, and broaden the applications of this classical problem. In Chapter 6, we present several representative application examples in communication networks and social-economic systems, aiming to illuminate the connections between the Bayesian and the frequentist formulations of bandit problems and how structural results pertaining to one may be leveraged to obtain solutions under the other.

KEYWORDS

multi-armed bandit, machine learning, online learning, reinforcement learning, Markov decision processes

To Peter Whittle
and to Lang and Everett.

Contents

Preface

The term "multi-armed bandit" comes from likening an archetypal online learning problem to playing a slot machine that has multiple arms (slot machines are also known as bandits due to their ability to empty the player's pockets). Each arm, when pulled, generates random rewards drawn from an unknown distribution or a known distribution with an unknown mean. The player chooses one arm to pull at each time, with the objective of accumulating, in expectation, as much reward as possible over a given time horizon. The tradeoff facing the player is a classic one, that is, to explore a less observed arm which may hold a greater potential for the future or to exploit an arm with a history of offering good rewards. It is this tension between learning and earning that lends complexity and richness to the bandit problems.

As in many problems involving unknowns, bandit problems can be treated within the Bayesian or frequentist frameworks, depending on whether the unknowns are viewed as random variables with known prior distributions or as deterministic quantities. These two schools have largely evolved independently. In recent years, we witness increased interests and much success in cross-pollination between the two schools. It is my hope that by covering both the Bayesian and frequentist bandit models, this book further stimulates research interests in this direction.

We start in Chapter 1 with an overview on the history and foundational results of the bandit problems within both frameworks. In Chapters 2 and 4, we devote our attention to the canonical Bayesian and frequentist formulations. Major results are treated in detail. Proofs for key theorems are provided.

New and emerging applications in computer science, engineering, and social-economic systems give rise to a diverse set of variants of the classical models, generating new directions and bringing in new techniques to this classical problem. We discuss major variants under the Bayesian framework and the frequentist framework in Chapters 3 and 5, respectively. The coverage, inevitably incomplete, focuses on the general formulations and major results with technical details often omitted. Special attention is given to the unique challenges and additional structures these variants bring to the original bandit models. Being derivative to the original models, these variants also offer a deeper appreciation and understanding of the core theory and techniques. In addition to bringing awareness of new bandit models and providing reference points, these two chapters point out unexplored directions and open questions.

In Chapter 6, we present application examples of the bandit models in communication networks and social-economic systems. While these examples provide only a glimpse of the expansive range of potential applications of bandit models, it is my hope that they illustrate two fruitful research directions: applications with additional structures that admit stronger results than what can be offered by the general theory, and applications bringing in new objectives and constraints that push the boundaries of the bandit models. These examples are chosen also to show the connections between the Bayesian and frequentist formulations and how structural results pertaining to one may be leveraged to obtain solutions under the other.

Qing Zhao
Ithaca, NY, August 2019

Acknowledgments

In December 2015, Srikant, the editor of the series, asked whether I would be interested in writing a book on multi-armed bandits. By that time, I had worked on bandit problems for a decade, starting with the Bayesian and then the frequentist. I was quite confident in taking on the task and excited with the ambition of bringing together the two schools of approaches together within one book, which I felt was lacking in the literature and was much needed. When asked of a timeframe for finishing the book, I gave an estimate of one year. "That ought to leave me plenty of margin." I thought. My son, Everett, was one year old then.

Everett is starting kindergarten next week.

Writing this book has been a humbling experience. The vast landscape of the existing literature, both classical and new, reincarnations of ideas, often decades apart, and quite a few reinventions of wheels (with contributions from myself in that regard), have made the original goal of giving a comprehensive coverage and respecting the historical roots of all results seem unattainable at times. If it were not for the encouragement and persistent nudging from Srikant and the publisher Michael Morgan, the book would have remained unfinished forever. I do not think I have achieved the original goal. This is a version I can, at least, live with.

Many people have helped me in learning this fascinating subject. The first paper I read on bandit problems was "Playing Golf with Two Balls" pointed to me by Vikram Krishnamurthy, then a professor at UBC and now my colleague at Cornell. It was the summer of 2005 when I visited Vikram. We were working on a sensor scheduling problem under the objective of network lifetime maximization, which leads to a stochastic shortest-path bandit. The appeal of the bandit problems was instantaneous and has never faded, and I must admit a stronger affection towards the Bayesian models, likely due to the earlier exposure. Special thanks go to Peter Whittle of the University of Cambridge. I am forever grateful to his tremendous encouragement through my career and generous comments on our results on restless bandits. His incisive writing has always been an inspiration. Many thanks to my students, past and current, who taught me most things I know about bandits through presentations in our endless group meetings and through their research, in particular, Keqin Liu and Sattar Vakili whose dissertations focused almost exclusively on bandit problems.

My deepest appreciation goes to my husband, Lang, for letting me hide away for weeks finishing up a first draft while he took care of Everett, for agreeing to read the draft and providing comments and actually did so for the Introduction! I thank my dear Everett for the many hours sitting patiently next to me, copying on his iPad every letter I typed. It was this July in Sweden when there was no daycare and I was trying to wrap up the book. He has been the most faithful reader of the book, who read, not word by word, but letter by letter.

I thank the U.S. National Science Foundation, the Army Research Office, and the Army Research Lab for supporting the research of my group and the talented Ph.D. students I worked with. I am grateful to the support during my 2018–2019 sabbatical leave in Sweden from the European Union through a Marie Skłodowska-Curie grant[1] and from the Chalmers University of Technology through a Jubilee Professorship. Their generous support has allowed me to put in continuous effort on the book and finally push it over the finish line.

Qing Zhao
Ithaca, NY, August 2019

[1]Current supporting grants include the National Science Foundation Grant CCF-1815559, the Army Research Laboratory Network Science CTA under Cooperative Agreement W911NF-09-2-0053, and the European Unions Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754412.