SpringerBriefs in Computer Science

For further volumes:
http://www.springer.com/series/10028

Richard Berk

Criminal Justice Forecasts of Risk

A Machine Learning Approach

Springer

Richard Berk
University of Pennsylvania
School of Arts & Sciences
Philadelphia, PA
berkr@sas.upenn.edu

ISSN 2191-5768 e-ISSN 2191-5776
ISBN 978-1-4614-3084-1 e-ISBN 978-1-4614-3085-8
DOI 10.1007/978-1-4614-3085-8
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012933586

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book is an effort to put in one place and in accessible form the most recent work on forecasting re-offending by individuals already in criminal justice custody. Much of that work is my own and comes from over two decades of close collaborations with a number of criminal justice agencies. After many requests to provide in one place an account of the procedures I have used, I agreed to write this book. What I hope distinguishes the material from what has come before is the use of machine learning statistical procedures coupled with very large datasets, an explicit introduction of the relative costs of forecasting errors as the forecasts are constructed, and an exclusive statistical focus on maximizing forecasting accuracy. Whether the forecasts that result are "good enough," I leave to the reader.

The audience for the book is graduate students and researchers in the social sciences, and data analysts in criminal justice agencies. Formal mathematics is used only as necessary or in concert with more intuitive explanations. A working knowledge of the generalized linear model is assumed. All of the empirical examples were constructed using the programming language R, in part because most of the key tools are not readily available in the usual social science point-and-click statistical packages. R runs on a wide variety of platforms and is available at no cost. It can be downloaded from www.r-project.org/.

A very large number of individuals have helped me as my criminal justice forecasting activities have evolved. On technical matters, I owe a special debt to David Freedman who put up with my questions over 10 years of collaboration. More recently, I have benefited enormously working the colleagues in Penn's Department of Statistics, especially Larry Brown, Andreas Buja, Ed George, Mikhail Traskin and Linda Zhao. On policy matters my mentor Peter Rossi provided a foundation on which I still draw. My hands-on education has come from a number of criminal justice officials starting with several talented individuals working for the California Department of Corrections in the 1990s: George Lehman, Penny O'Daniel and Maureen Tristan. More recently, Ellen Kurtz of Philadelphia's Adult Department of Probation and Parole was an early supporter and adopter modern forecasting methods from whom I learned a lot. Catherine McVey, chairman of the Pennsylvania Board of Probation and Parole played a key role in educating me about the parole decision

process and the *real politik* in which parole decisions are made. Jim Alibirio, as a key IT and data analyst person working for the parole board, was a true data maven and wonderful colleague as we together developed the required forecasting procedures. Other practitioners to whom I am especially indebted are Mark Bergstrom Executive Director of the Pennsylvania Sentencing Commission, and Sarah Hart, a Deputy Prosecutor (and much more) in the Philadelphia district attorney's office. A sincere thanks to one and all.

All of the analyses reported in the book would have been impossible without the data on which they are based. Collecting those data required hard work by a number of individuals from several criminal justice agencies at state and local levels. An important subset of the work was supported by a grant to the State of Pennsylvania from the National Institute of Justice — "Projecting Violent Re-Offending in a Parole Population: Developing A Real-Time Forecasting Procedure to Inform Parole Decision-Making" (# 2009-IJ-CX-0044). My most sincere thanks to NIJ for the funding and to Patrick Clark, who is a knowledgeable and constructive grant monitor with a great sense of humor.

Finally, I am a notoriously bad proof reader and can always use help with the quality of my arguments and prose. Justin Bleich and Anna Kegler carefully read an earlier draft of this manuscript and provided all manner of assistance. However, the responsibility for any remaining typos or unclear prose is surely mine.

Contents

1 Getting Started 1
1.1 Setting the Stage 1
1.1.1 A Brief Motivating Example 4

2 Some Important Background Material 7
2.1 Policy Considerations 7
2.1.1 Criminal Justice Forecasting Goals 7
2.1.2 Decisions to be Informed by the Forecasts 9
2.1.3 Outcomes To be Forecast 10
2.2 Data Considerations 11
2.2.1 Stability Over Time 11
2.2.2 Training Data and Test Data 12
2.2.3 Representative Data 13
2.2.4 Large Samples 14
2.2.5 Data Continuity 15
2.2.6 Missing Data 16
2.3 Statistical Considerations 17
2.3.1 Actuarial Methods 18
2.3.2 Building in the Relative Costs of Forecasting Errors 19
2.3.3 Effective Forecasting Algorithms 22

3 A Conceptual Introduction to Classification and Forecasting 27
3.1 Populations and Samples 27
3.2 Classification and Forecasting Using Decision Boundaries 27
3.3 Classification by Data Partitions 30
3.4 Forecasting by Data Partitions 33
3.5 Finding Good Data Partitions 34
3.6 Enter Asymmetric Costs 35
3.7 Recursive Partitioning Classification Trees 37
3.7.1 How Many Terminal Nodes? 39
3.7.2 Classification Tree Instability 40

4 A More Formal Treatment of Classification and Forecasting 43
4.1 Introduction 43
4.2 Data Generation Models 44
4.3 Notation 45
4.4 Classification 45
4.5 Estimation in the Real World 47
4.5.1 Estimation Bias 48
4.5.2 Estimation Variance 50
4.5.3 The Bias-Variance Tradeoff 51
4.5.4 Uncertainty 53
4.6 A Bit More on the Joint Probability Model 54
4.7 Summary 56

5 Tree-Based Forecasting Methods 59
5.1 Introduction 59
5.2 Splitting the Data 60
5.3 Building the Costs of Classification Errors 62
5.4 Classification Tables 63
5.5 Ensembles of Trees: Random Forests 64
5.5.1 Variable Importance 66
5.5.2 Response Functions 68
5.5.3 Forecasting 69
5.5.4 Procedure Uncertainty 70
5.5.5 Forecast Uncertainty 71
5.6 Tree-Based Alternatives to Random Forests 73
5.6.1 Stochastic Gradient Boosting 73
5.6.2 Bayesian Trees 75
5.7 Why Do Ensembles of Classification Trees Work So Well? 77

6 Examples 81
6.1 A Simplified Example 81
6.2 A More Complex Example 88

7 Implementation 101
7.1 Making it Real 101
7.2 Demonstration Phase 101
7.3 Hardware and Software 102
7.4 Data Preparation for Forecasting 103
7.5 Personnel 104
7.6 Public Relations 105

8 Some Concluding Observations About Actuarial Justice and More 107
8.1 Current Trends in Actuarial Justice 107
8.2 Larger Issues 109
8.2.1 Forecasts Despite Individual Uniqueness 109
8.2.2 Risk and Forecasting 110

8.2.3 Black-Box Methods 110
8.3 Dark Structure and New Science 111
8.4 A Closing Sermonette 112

References 113