Coding Ockham's Razor

Lloyd Allison

Coding Ockham's Razor

Springer

Lloyd Allison
Faculty of Information Technology
Monash University
Melbourne, Victoria, Australia

ISBN 978-3-319-76432-0 ISBN 978-3-319-76433-7 (eBook)
https://doi.org/10.1007/978-3-319-76433-7

Library of Congress Control Number: 2018936916

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Sally, Bridget, Jean, Yeshi, Nyima, and Lhamo.

Preface

The minimum message length (MML) principle was devised by Chris Wallace (1933–2004) and David Boulton in the late 1960s [12, 93] initially to solve the unsupervised mixture modelling problem—an important problem, a mathematical analysis, and a working computer program (Snob) that gives useful results in many different areas of science, a "complete" research project.

The Foundation Chair of Computer Science at Monash University, Chris is also particularly remembered for his work on the "Wallace multiplier" [85, 86], pseudo-random number generators [14, 89], and operating systems [6, 99].

MML was developed [91, 92] in practical and theoretical directions and was applied to many inference problems by Chris, co-workers, postgraduates, and postdocs. One of my personal favourite applications is Jon Patrick's modelling of megalithic stone circles [65, 66].

I first heard about MML over lunch one day which led to applying it to biological sequence alignment [3] and related problems [15], and eventually after many twists and turns to protein structural alignment [17] and protein folding patterns [83].

Unfortunately much MML-based research that led to new inductive inference programs resulted in little shared software *componentry*. A new program tended to be written largely from scratch by a postgrad, postdoc or other researcher and did not contribute to any software library of shared parts. As such the programs embody reimplementations of standard parts. This phenomenon is not due to any special property of MML and actually seems to be quite common in research but it is rather ironic because, what with the complexity of models and of data being measured in the same units, MML is well suited to writing components that can be reused and supplied as parameters to other inductive inference software.

The first MML book is the one written by Chris Wallace [92] but published posthumously; it is the reference work for MML theory. This other MML book is an attempt to do a combined MML and Software Engineering analysis of inductive inference software. Sound programming skills are needed to write new application programs for inductive inference problems. Some mathematical skills, particularly in calculus and linear algebra, are needed to do a new MML analysis of one's favourite statistical model.

Melbourne, Victoria, Australia
Lloyd Allison

Acknowledgements

Chris Wallace was a great inspiration and always generous with ideas. He is sadly missed. This book was begun largely at the urging of Arun Konagurthu who at times shows aspects of both an irresistible force and an immovable object. He also contributed to the content and examples and fought valiantly in the typesetting wars.

Leigh Fitzgibbon, Josh Comley, and Rodney O'Donnell deserve special mention for contributions [18, 34, 61] to early attempts to create and use *general* MML software.

Many thanks go to Dianna Kenny for sharing her data on musicians and mortality (Sect. 7.5).

I am indebted to those who read parts, a little or a lot, of drafts of the book and who suggested improvements, a few or many, in alphabetical order: Rohan Baxter, Minh Duc Cao, Trevor Dix, Rodney O'Donnell, Arun Konagurthu, Francois Petitjean, Joel Reicher, Daniel Schmidt. But, as they say, the mistakes are all my own.

Contents