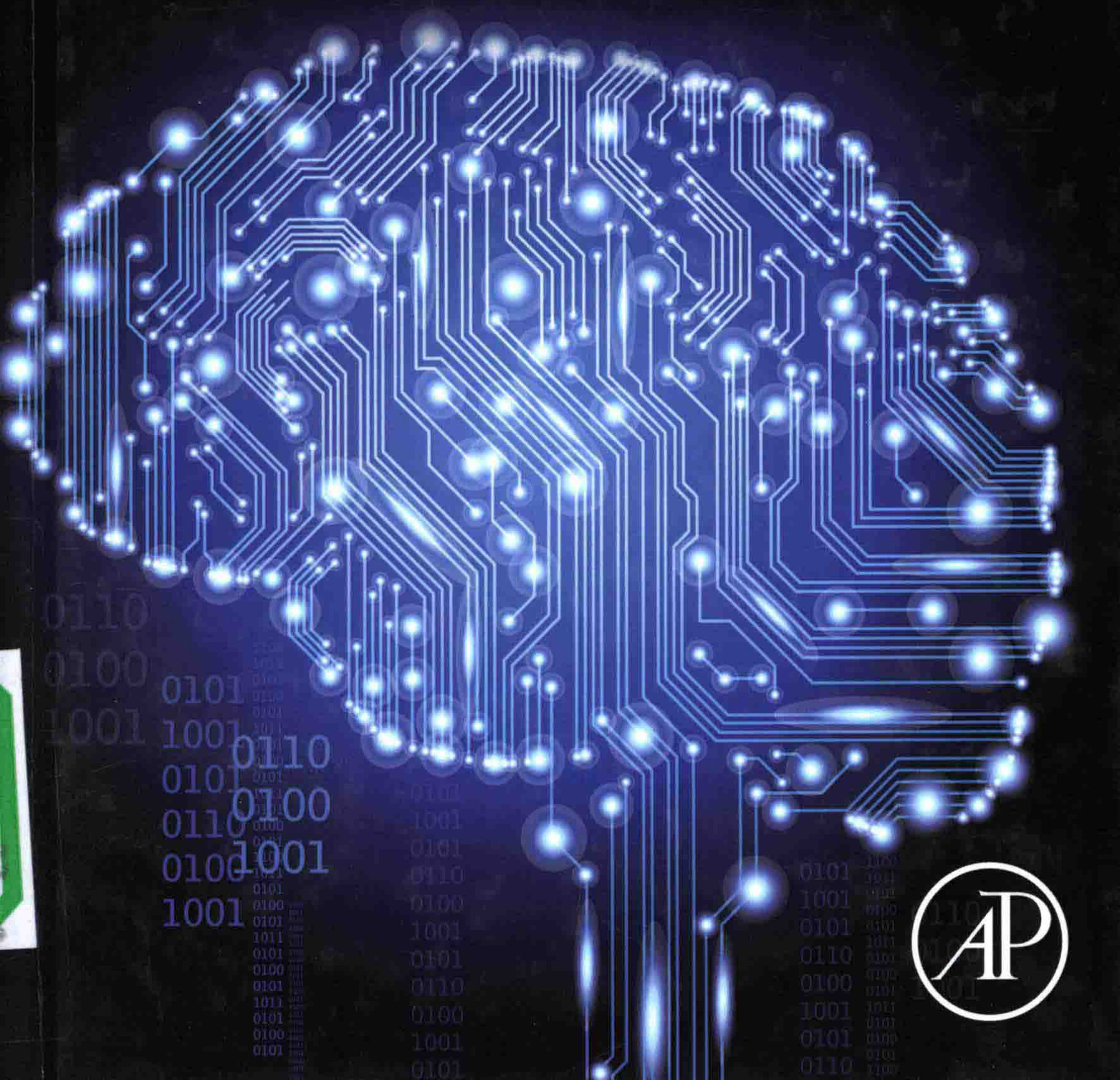
CALCULUS OF [redacted]UGHT

NEUROMORPHIC LOGISTIC REGRESSION IN COGNITIVE MACHINES

DANIEL M. RICE



Calculus of Thought: Neuromorphic Logistic Regression in Cognitive Machines

DANIEL M. RICE

ELSEVIER

Amsterdam • Boston • Heidelberg • London
New York • Oxford • Paris • San Diego
San Francisco • Sydney • Tokyo
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
225, Wyman Street, Waltham, MA 02451, USA
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-12-410407-5

For information on all Academic Press publications
visit our web site at store.elsevier.com

Printed and bound in USA
14 15 16 17 10 9 8 7 6 5 4 3 2 1

Calculus of Thought: Neuromorphic Logistic Regression in Cognitive Machines

DEDICATION

To the memory of my father Richard Rice

PREFACE

A PERSONAL PERSPECTIVE

I first heard the phrase Calculus of Thought as a second year graduate student in 1981 when I took a seminar called *Brain and Behavior* taught by our professor James Davis at the University of New Hampshire. He gave us a thorough grounding in the cognitive neuroscience of that day, and he spent significant time on various models of neural computation. The goal of cognitive research in neuroscience, he constantly reminded us, was to discover the neural calculus, which he took to be a complete and unifying understanding of how the brain performs computation in diverse functions such as coordinated motor action, perception, learning and memory, decision making, and causal reasoning. He suggested that if we had such a calculus, then there would be truly amazing practical artificial intelligence applications beyond our wildest dreams. This was before the widespread popularity of artificial neural network methods in the mid-1980s, as none of the quantitative models that we learned about were artificial neural networks, but instead were grounded in empirical data in neural systems like the primary visual cortex and the basal ganglia. He admitted that the models that we were taught fell way short of such a calculus, but he did fuel an idea within me that has stayed for more than 30 years.

I went on to do my doctoral dissertation with my thesis advisor Earl Hagstrom on timing relationships in auditory attention processing as reflected by the scalp recorded Electroencephalography (EEG) alpha rhythm. During the early and mid-1980s, there was not a lot of interest in the EEG as a window into normal human cognition, as the prevailing sentiment was that the brain's electric field potentials were too gross of a measure to contain valuable information. We now have abundant evidence that brain field potential measures, such as the EEG alpha rhythm, are sensitive to oscillating and synchronous neural signals that do reflect the rhythm and time course of cognitive processing. My dissertation study was one of a handful of initial studies to document a parallelism between the oscillating neural synchrony as measured by the EEG and cognition.[1] Through that dissertation study, and the advice that I also got from other faculty committee members—John Limber and Rebecca Warner, I learned the importance of well-controlled experimental findings in providing more reliable explanations. Yet, this dissertation did not address how basic cognitive computations might be performed by oscillating neural synchrony mechanisms, and my thoughts have wandered back to how this Calculus of Thought might work ever since.

I did postdoctoral fellowship research with Monte Buchsbaum at the University of California-Irvine Brain Imaging Center. Much of our work focused on abnormal temporal lobe EEG slowing in probable Alzheimer's patients and related temporal lobe slowing in nondemented older adults. We published one paper in 1990 that replicated other studies including one from Monte's lab showing abnormal temporal lobe EEG slowing in Alzheimer's patients compared to nondemented elderly, but we critically refined the methodology to get a higher fidelity measurement.[2] We published a second paper in 1991 that used this refined methodology to make the first claim that Alzheimer's disease must have an average preclinical period of at least 10 years.[3] Our basic finding was that a milder form of the temporal lobe EEG slowing observed in Alzheimer's patients was seen in nondemented older adults with minor memory impairment. This observed memory impairment had the exact same profile as what neuropsychologist Brenda Milner had observed in the famous medial temporal lobe patient H.M. This is the most famous clinical case study in neuroscience.[4] This was that there was normal immediate recall ability, but dramatically greater forgetting a short time later. Other than this memory deficit, we could find nothing else wrong in terms of cognitive and intelligence tests with these nondemented older adults. Given the nature of the memory deficit and the location of the EEG abnormality, we suggested that the focus of this abnormality must be in the medial temporal lobe of the brain. Given the prevalence of such EEG slowing in nondemented older adults and the prevalence of Alzheimer's disease, we calculated that this must be a preclinical sign of Alzheimer's disease that is present 10 years earlier than the clinical diagnosis. We had no idea how abnormal neural synchrony might be related to dysfunction in memory computations, but ever since my thoughts have wandered back to how this Calculus of Thought could go awry early in Alzheimer's disease.

Our claim that there is a long preclinical period in Alzheimer's disease with a major focal abnormality originating in the medial temporal lobe memory system has now become the prevailing view in Alzheimer's research.[5,6,7] This evidence for the long preclinical period is such that the National Institute on Aging issued new diagnostic recommendations in

2011 to incorporate preclinical Alzheimer's disease into the clinical Alzheimer's disease diagnosis.[8] Yet, when I moved to a new assistant professor position at the University of Southern California (USC) in the earlier 1990s and tried to get National Institutes of Health (NIH) funding to validate this hypothesis with a longitudinal, prospective study, my proposals were repeatedly not funded. It was not controversial that Alzheimer's starts with medial temporal lobe memory system problems, as almost everyone believed that was true. However, I was a complete newcomer, and the idea that Alzheimer's had such a long preclinical period was just too unexpected to be accepted as reasonable by the quorum needed to get NIH funding. It was then that I began to have recollections of Thomas Kuhn's theory about bias in science taught by my undergraduate history of science professor Charles Bonwell many years earlier. Kuhn warned that the established scientific community is often overinvested in the basic theory in their field that he called "paradigm".[9] Kuhn argued that this bias precludes most scientists from considering new results and hypotheses that are inconsistent with the paradigm. In the early 1990s, the entire Alzheimer's clinical diagnostic process and pharmaceutical research were only concerned with actual demented patients. So any results or explanation that there is in fact a long 10-year preclinical period was not yet welcome. As it turns out, the only hope to prevent and treat Alzheimer's disease now seems to be in this very early period when the disease is still mild,[10] as all drug studies in actual demented patients have not really worked. At that time, I was not yet familiar with Leibniz's concept of *Calculus Ratiocinator*, or a thought calculus machine designed to generate explanations that avoid such bias. Though, I did start to think that any unbiased machine that could generate most reasonable hypotheses based upon all available data would be useful.

Fortunately, I had wider research interests than Alzheimer's disease, as I had an interest in quantitative modeling of the brain's electric activity as a means to understand the brain's computations. One day while still at UC-Irvine, I attended a seminar given by a graduate student on maximum entropy and information theory in a group organized by mathematical cognitive psychologists Duncan Luce and Bill Batchelder. I then began to study maximum entropy on my own and became interested in the possibility that this could be the basic computational process within neurons and the brain. When I ended up teaching at the USC a few years later, I was fortunate enough to collaborate with engineering professor Manbir Singh and his graduate student Deepak Khosla on modeling the EEG with the maximum entropy method.[11] In our maximum entropy modeling, Deepak taught me a very interesting new way to smooth error out of EEG models using what we now call L2 norm regularization. Yet, I also began to think that there might be a better approach based upon probability theory to model and ultimately reduce error in regression models that model the brain's neural computation. This thinking eventually led to the reduced error logistic regression (RELR) method that is the proposed Calculus of Thought, which is the subject of this book.

In April of 1992, I had a bird's eye view of the Los Angeles (LA) riots through my third floor laboratory windows at USC that faced the south central section of the city. I watched shops and houses burn, and I was shocked by the magnitude of the violence. But, I also began to wonder whether human social behavior also might be determined probabilistically in ways similar to how causal mechanisms determine cognitive neural processes like attention and memory so that it might be possible to predict and explain such behavior. After the riots, I listened to the heated debates about causal forces involved in the 1992 LA riots, and again I began to wonder how objective these hypotheses about causal explanations of human behavior ever could be due to extremely strong biases. This was also true of most explanations of human behavior that I saw offered in social science whether they were conservative or liberal. So, it became clear to me that bias was the most significant problem in social science predictions and explanations of human behavior. And, I began to believe that an unbiased machine learning methodology would be a huge benefit to the understanding of human behavior outcomes. However, I did not yet make the connection that a data-driven quantitative methodology that models neural computations could be the basis of this unbiased *Calculus Ratiocinator* machine.

Eventually, it became clear that my proposed research on a longitudinal, prospective study to test the putative long preclinical period in Alzheimer's disease would not be funded. So, because NIH funding was required to get tenure at USC, I realized that I had better find another career. I resigned from USC in 1995, which was years prior to when my tenure decision would have been made. Even though I had a contract that would be renewed automatically until at least 1999, I saw no reason to be dead weight for several years and then face a negative tenure decision. Instead, I wanted to get started on a more entrepreneurial career where I would have more control over my fate. So, I pursued an applied analytics career. I have enjoyed this career often better than my academic career because the rewards and feedback are more immediate. Also, there is much more emphasis on problem solving as the goal rather than writing a paper or getting a grant, although sales ability is obviously always valued. Yet, I eventually realized that the same bias problems that limit understanding of human behavior in academic neural, behavioral, and social science also limit practical business analytics. Time and time again I noticed decisions based upon biased predictive or explanatory models dictated by incorrect, questionable assumptions, so again I noticed a need for a data-driven *Calculus Ratiocinator* to generate reasonable and testable hypotheses and conclusions that avoid human bias.

Much of my applied analytic work has involved logistic regression, and I eventually learned that maximum entropy and maximum likelihood methods yield identical results in modeling the kind of binary signals that both neurons and business decisions produce. Thus, at some point I also realized that I could continue my research into the RELR Calculus of Thought. But instead of focusing on the brain's computations, I could focus on practical real world business applications using the same computational method that I believed that neurons use—the RELR method. In so doing, I eventually realized that RELR could be useful as the *Calculus Ratiocinator* that Leibniz had suggested is needed to remove biased answers to important, real world questions.

Many of today's "data scientists" have a background in statistics, pure mathematics, computer science, physics, engineering, and operations research. Yet, these are academic areas that are not focused on studying human behavior, but instead focus on quantitative and technical issues related to more mechanical data processes. Many other analytic scientists, along with many analytic executives, have a background in behaviorally oriented fields like economics, marketing, business, psychology, and sociology. These academic areas do focus on studying human behavior, but are not heavily quantitatively oriented. There is a real need for a theoretical bridge between the more quantitative and more behavioral knowledge areas, and that is the intention of this book. So, I believe that this book could appeal both to quantitative and behaviorally oriented analytic professionals and executives and yet fill important knowledge gaps in each case.

Unless an analytic professional today has a specific background in cognitive or computational neuroscience, it is unlikely that they will have a very good understanding of neuroscience and how the brain may compute cognition and behavior. Yet, data-driven modeling that seeks to solve the Turing problem and mimic the performance of the human brain without its propensity for bias and error is probably the implicit goal of all machine-learning applications today. In fact, this book argues that cognitive machines will need to be neuromorphic, or based upon neuroscience, in order to simulate aspects of human cognition. So, this book covers the most fundamental and important concepts in modern cognitive neuroscience including neural dynamics, implicit and explicit learning, neural synchrony, Hebbian spike-timing dependent plasticity, and neural Darwinism. Although this book presents a unified view of these fundamental neuroscience concepts in relationship to RELR computation mechanisms, it also should serve as a much needed introductory overview of these fundamental cognitive neuroscience principles for analytic professionals who are interested in neuromorphic cognitive machines.

Theories in science can simplify a subject and make that subject more understandable and applicable in practice, even when theories ultimately are modified with new data. So this book may help analytic professionals understand fundamental neural computational theory and cognitive neuroscientists understand practical issues surrounding real world machine learning applications. It is true that the theory of the brain's computation in this book should be viewed as a question rather than an answer. But, if cognitive neuroscientists and analytic professionals all start asking questions about whether the brain and machine learning each can work according to the information theory principles that are laid out in this book, then this book will have served its purpose.

Daniel M. Rice
St Louis, MO (USA)
Autumn, 2013

CONTENTS

CHAPTER 1

Calculus Ratiocinator

"It is obvious that if we could find characters or signs suited for expressing all our thoughts as clearly and as exactly as arithmetic expresses numbers or geometry expresses lines, we could do in all matters insofar as they are subject to reasoning *all that we can do in arithmetic and geometry. For all investigations which depend on reasoning would be carried out by transposing these characters and by a species of calculus."*

Gottfried Leibniz, Preface to the General Science, 1677.[1]

Contents

At the of end of his life and starting in 1703 Gottfried Leibniz engaged in a 12-year feud with Isaac Newton over who first invented the calculus and who committed plagiarism. All serious scholarship now indicates that both Newton and Leibniz developed calculus independently.[2] Yet, stories about Leibniz's invention of calculus usually focus on this priority dispute with Newton and give much less attention to how Leibniz's vision of calculus differed substantially from Newton's. Whereas Newton was trained in mathematical physics and continued to be associated with academia during the most creative time in his career, Leibniz's early academic failings in math led him to become a lawyer by training and an entrepreneur by profession.[3] So Leibniz's deep mathematical insights that led to calculus occurred away from a university professional association. Unlike Newton whose entire mathematical interests seemed tied to physics, Leibniz clearly had a much broader goal for calculus. These applications were in areas well beyond physics that seem to have nothing to do with mathematics. His dream application was for a *Calculus Ratiocinator*, which is synonymous with Calculus of Thought.[4] This can be interpreted to be a very precise mathematical model of cognition that could be automated in a machine to

Calculus of Thought
http://dx.doi.org/10.1016/B978-0-12-410407-5.00001-5

answer any important philosophical, scientific, or practical question that traditionally would be answered with human subjective conjecture.[5] Leibniz proposed that if we had such a cognitive calculus, we could just say "Let us calculate"[6] and always find most reasonable answers uncontaminated by human bias.

In a sense, this concept of *Calculus Ratiocinator* foreshadows today's predictive analytic technology.[7] Predictive analytics are widely used today to generate better than chance longer term projections for more stable physical and biological outcomes like climate change, schizophrenia, Parkinson's disease, Alzheimer's disease, diabetes, cancer, optimal crop yields, and even good short-term projections for less stable social outcomes like marriage satisfaction, divorce, successful parenting, crime, successful businesses, satisfied customers, great employees, successful ad campaigns, stock price changes, loan decisions, among many others. Until the widespread practice of predictive analytics with the introduction of the computers in the past century, most of these outcomes were thought to be too capricious to have anything to do with mathematics. Instead, they were traditionally answered with speculative and biased hypotheses or intuitions often rooted in culture or philosophy (Fig. 1.1).

Figure 1.1 Gottfried Wilhelm Leibniz.[8]

Until just very recently, standard computer technology could only evaluate a small number of predictive features and observations. But, we are now in an era of big data and high performance massively parallel computing. So our predictive models should now become much more powerful. This is because it would seem reasonable that those traditional methods that worked to select important predictive features from small data will scale to high-dimension data and suddenly select predictive models that are much more accurate and insightful. This would give us a new and much more powerful big data machine intelligence technology that is everything that Leibniz imagined in a *Calculus Ratiocinator.* Big data massively parallel technology should thus theoretically allow completely new data-driven cognitive machines to predict and explain capricious outcomes in science, medicine, business, and government.

Unfortunately, it is not this simple. This is because observation samples are still fairly small in most of today's predictive analytic applications. One reason is that most real-world data are not representative samples of the population to which one wishes to generalize. For example, the people who visit Facebook or search on Google might not be a good representative sample of many populations, so smaller representative samples will need to be taken if the analytics are to generalize very well. Another problem is that many real-world data are not independent observations and instead are often repeated observations from the same individuals. For this reason, data also need to be down sampled significantly to be independent observations. Still, another problem is that even when there are many millions of independent representative observations, there are usually a much smaller number of individuals who do things like respond to a particular type of cancer drug or commit fraud or respond to an advertising promotion in the recent past. The informative sample for a predictive model is the group of targeted individuals and a group of similar size that did not show such a response, but these are not usually big data samples in terms of large numbers of observations. So, the biggest limitation of big data in the sense of a large number of observations is that most real-world data are not "big" and instead have limited numbers of observations. This is especially true because most predictive models are not built from Facebook or Google data.[9]

Still, most real-world data are "big" in another sense. This is in the sense of being very high dimensional given that interactions between variables and nonlinear effects are also predictive features. Previously we have not had the technology to evaluate high dimensions of potentially predictive variables rapidly enough to be useful. The slower processing that was the reason for this "curse of dimensionality" is now behind us. So many might believe that this suddenly allows the evaluation of almost unfathomably high dimensions of data for the selection of important features in much more accurate and smarter big data predictive models simply by applying traditional widely used methods.

Unfortunately, the traditional widely used methods often do not give unbiased or non-arbitrary predictions and explanations, and this problem will become ever more apparent with today's high-dimension data.

1. A FUNDAMENTAL PROBLEM WITH THE WIDELY USED METHODS

There is one glaring problem with today's widely used predictive analytic methods that stands in the way of our new data-driven science. This problem is inconsistent with Leibniz's idea of an automated machine that can reproduce the very computations of human cognition, but without the subjective biases of humans. This problem is suggested by the fact that there are probably at least hundreds of predictive analytic methods that are in use today. Each method makes differing assumptions that would not be agreed upon by all, and all have at least one and sometimes many arbitrary parameters. This arbitrary diversity is defended by those who believe a "no free lunch" theorem that argues that there is no one best method across all situations.[10,11] Yet, when predictive modelers test various arbitrary algorithms based upon these methods to get a best model for a specific situation, they obviously will only test but a tiny subset of the possibilities. So unless there is an obvious very simple best model, different modelers will almost always produce substantially different arbitrary models with the same data.

As examples of this problem of arbitrary methods, there are different types of decision tree methods like CHAID and CART which have different statistical tests to determine branching. Even with the very same method, different user-provided parameters for splitting the branches of the tree will often give quite different decision trees that will generate very different predictions and explanations. Likewise, there are many widely used regression variable selection methods like stepwise and LASSO logistic regression that are all different in the arbitrary assumptions and parameters employed in how one selects important "explanatory" variables. Even with the very same regression method, different user choices in these parameters will almost always generate widely differing explanations and often substantially differing predictions. There are other methods like Principal Component Analysis (PCA), Variable Clustering and Factor Analysis that attempt to avoid the