The Art and Science of Machine Intelligence

Walker H. Land Jr. • J. David Schaffer

The Art and Science of Machine Intelligence

With An Innovative Application for Alzheimer's Detection from Speech

Springer

Walker H. Land Jr.
Binghamton University
Bowie, MD, USA

J. David Schaffer
Binghamton University
Binghamton, NY, USA

ISBN 978-3-030-18495-7 ISBN 978-3-030-18496-4 (eBook)
https://doi.org/10.1007/978-3-030-18496-4

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

Foreword

Today, we live in a hyper-connected, technology-immersed environment enabled by foundational architectures with key interdependent components, such as big data, cloud computing, mobility, and cybersecurity. Machine intelligence (learning) is an advanced capability that can be leveraged to help optimize these components as well as the application and data layers above. I believe effectively advancing the art and science of machine intelligence, as described in this book, will have a positive impact on improving accuracy and performance of decision-based systems across multiple domains.

Upon submission of the book manuscripts for publication, the authors recommended based on my relevant experience and book review efforts that I would be a suitable author for this foreword. My background includes design/development of data analytics solutions, executive leadership of company-wide research development portfolios, and currently serving as the Capitol Technology University Board Chairman, where data science and cybersecurity doctoral programs exist, and serving on the Dean Leadership Council at Syracuse University Engineering and Computer Science College, where engineering and computer science doctoral programs exist.

I know Walker as my admired father and David as a professional colleague. I have often referenced their intellect, integrity, and/or kindness to peers. Walker's proudest technical achievement was successfully developing the trajectories for the Saturn Apollo lunar mission and has published one other book. He also feels blessed to see the successes achieved by both his children and grandchildren. David's proudest technical achievement was taking evolutionary computation into the domain of multiobjective selection and has 43 patents. They have worked well together for the last 13 years predominately in the areas of machine learning and pattern discovery. Each of them experienced accomplished lives and are now focused on their particular passion to help make this world a better place. I refer the reader to the publishers online author biographies for additional information.

In a consistent manner, this book reinforces the intrinsic value of creative thinking supported by sound problem-solving using existing, refined, and new methods of machine learning. These methods are explained in detail through judicious application of novel combinations of pattern discovery while placing laser focus on demonstrating that techniques are most successfully applied when both intelligence processing (the science part) is combined with the training (the art part) of very complex machine intelligence designs. The majority of the supporting case studies are focused on innovative ways for detecting Alzheimer's; however, the authors do a credible job explaining how the methods used are domain independent. Furthermore, a majority of chapters have unique problem sets, varying in difficulty, for the reader to routinely apply and test comprehension.

The technical scope ranges from the introduction of key foundational concepts to the application of complex theory for solving pattern classification problems. I have read other books that do one or the other well but have not come across one that does both well in a unique and convincing way. The uniqueness of this book is the way it presents a number of approaches to aspects of machine pattern discovery, some invented by the authors and published separately in papers that they have integrated for the first time in one place. It effectively illustrates these diverse elements being applied in combinations to achieve sound results. I was impressed by the effective utilization of theory and application of mathematics throughout the work. The authors are consistent in the use of static pattern discovery tasks where the datasets are of limited size. An important takeaway for me—a practitioner of data analytics does not necessarily require mounds of data to determine an optimal answer when using these integrated methods.

The book's opening chapters detail genetic algorithms (GA), support vector machines (SVM), and generalized regression neural networks (GRNN). Alzheimer's disease (AD) pathology, AD effects on speech, and associated analysis follow to help understand the case domain. Complexity then steadily increases with GA/SVM paradigm, Bayesian networks (BN), Bayesian probabilistic neural networks (BPNN), mixture of experts/networks, and quantifying uncertainty.

With a combined 58 years of professional experience and 56 years of teaching experience, the authors' style and approach for the content has sufficient versatility to serve as a technical reference for the advanced professional or course textbook for students having the requisite foundational knowledge. I feel confident that the book can be effectively utilized for advanced undergraduate and graduate level courses as well as a useful reference to identify opportunities for doctoral research efforts. The authors identify a dozen opportunities for extending research throughout this book but the one that caught my attention was moving into dynamic pattern discovery (spatiotemporal patterns).

This book represents a scope of work by two authors of nearly 450 published artifacts that will motivate the sound-minded, well-versed engineer and the out-of-the-box research thinker to advance a specialty area within data analytics. There is a unique balance of risk-taking bounded by conservative validation thereby connecting left and right brain processing. For readers desiring an innovative,

informative, and applied deep dive in the how-to of pattern classification using machine learning then look no further than the Art and Science of Machine Intelligence. While the authors endeavor to provide a very thorough intuition for the operations described, to fully comprehend them a significant grasp of mathematics knowledge is recommended. I hope you enjoy and benefit from reading the content, I know I did.

Bowie, MD, USA Haden A. Land

Preface

In about 2006, the authors met in the Bioengineering Department at Binghamton University. They had previously worked independently in industry and academia, each with an interest in machine learning and pattern discovery, and each having pioneered a number of methods. Upon meeting they discovered that a number of methods previously developed and applied independently could be combined to good effect. Thus, began a very fruitful collaboration. By about 2017, they realized that together they had produced a body of work that might find fairly wide appeal, and so set about to write this book.

The application domain for these methods may be called the static pattern recognition problem, and as such includes a great many of the applications typically undertaken as machine learning. This task often begins with a question such as "do you suppose we might be able to detect that a subject has disease y (which we cannot observe directly) if we measure several features, x (which we can observe)?" A data set for such a task typically starts with a data array with a row for each subject and a column for each measurement (feature). Much activity focused today on "big data," frequently obtained from human use of the internet, is of this type. But also of this type are medical informatics tasks where the number of subjects is quite modest, usually because of the great difficulty and cost of acquiring volunteer subjects, and verifying the presence or absence of disease. The matrix rows are often divisible into two or more distinct groups, called classes, so the task is a classification task. The binary (two-class) tasks are the most frequent, but multi-class tasks are also quite prevalent. There is also the type of task where the "class" variable is not discrete, but a continuous variable. Perhaps the most venerable approach to tasks of this sort is linear regression. Often, there is little theoretical insight into the causal mechanisms that might drive the associations of the features and the outcome, so these efforts are sometimes called "fishing expeditions." Besides potentially producing a practical diagnostic instrument, the findings may also help focus efforts to acquire the understanding of the causal mechanisms. One such task, the detection of Alzheimer's disease using a sample of a person's speech is described in some detail and used as one illustration of the methods provided.

One task that is often a daunting first step in such research is the feature subset selection task. Some features may be redundant, correlated, or actually useless, but which ones? A method involving a combination of an evolutionary computation/genetic algorithm (GA) specifically devised for this task combined with the venerable support vector machine (SVM) learning classifier algorithm is described in the first few chapters. Another challenge may involve the best way to combine the predictions from several available classifiers, sometimes called ensemble of "mixture of experts" methods. One such method, the GRNN Oracle is described in Chap. 3. It is a maximum likelihood, minimum variance, unbiased estimator. This leads to the triumvirate GA-SVM-Oracle hybrid.

The next three chapters introduce Bayes theorem and Bayes networks. First, a basic introduction to the concepts and the challenges of building BNs from data. Then, the Bayesian probabilistic neural network (BPNN), a relative of the GRNN Oracle, is introduced as one way to evade some of the challenges of building BNs from data. Chapter 8 introduces another approach involving a simple three-layer BN design that seems appropriate for many diagnostic tasks and shows how to compute conditional probabilities that requires many fewer data than the traditional approaches.

The final chapter introduces a new method devised by the authors that we call teaching a learning classifier system to "know what it doesn't know" enabling a diagnostic system to flag predictions it knows may be unreliable. This then provides another possible metric, the uncertainty area, for comparing the performance of different learning classifier systems on the same data set.

Each method is introduced with the mathematical underpinning, but we have also attempted to present it intuitively, so readers not prepared to spend the time on the rigorous math may still acquire sufficient understanding to confidently apply it. In many cases, we provide pseudo computer code, so those with the necessary programming skills may write their own software. We also point to several open source software resources when they are available.

While some of the methods are well known, we believe this book is the first to bring together several others that were published in separate journal or conference papers, and most importantly to show the benefits of specific combinations.

Along the way we introduce what we believe is a novel method for normalizing measurement data, and an extension to the well-known area under the ROC curve that helps improve performance of these hybrid methods.

This book is intended for the following audiences: students, professionals, and researchers in the following areas: prediction, classification, detection, and diagnosis for the following fields: machine intelligence, complex adaptive systems, bioinformatics, and biomedical engineering. It is recommended for graduate course levels as well as some advanced undergraduate programs. The authors hope that it will also motivate researchers working in the above areas and fields to advance the technologies beyond the point discussed in this book.

Bowie, MD, USA — Walker H. Land Jr.
Binghamton, NY, USA — J. David Schaffer

Acknowledgements

We gratefully acknowledge the many creative colleagues we have had the privilege to know and work with. Larry Eshelman developed the initial CHC algorithm and continued to work on the subset selection task with Keith Mathias and Schaffer, leading to the MMX-SSS crossover operator. Timothy Masters and Suchindram Maniccam worked with Land to develop the original GRNN oracle. Stephen Zahorian and Roozbeh Sadeghian developed a new algorithm for the automatic speech recognition used to process the Alzheimer's speech samples. Kim Evanoski coordinated the social work graduate students who collected the speech samples. We thank the SUNY Research Foundation for a seed grant that made the whole effort possible, and the continuing support of Laura Bronstein, Dean of the Binghamton University College of Community and Public Affairs. We gratefully acknowledge the generosity of Brian Roark, Meg Mitchell, and Kristy Hollingshead who shared their speech processing software tools. Joseph Y. Lo worked with Land and published many papers using the breast cancer dataset developed at DUKE University. We also acknowledge many students, both graduates and undergraduates who contributed in ways too numerous to list. Finally, we acknowledge the contributions of Haden Land in reading the chapter drafts, suggesting editorial improvements and that problem sets be added, writing the Foreword and soliciting some among his extensive network for capable colleagues who were willing to draft endorsements. In addition, Professor Land acknowledges Haden for keeping Dr. Schaffer informed on the status and outcome of a critical brain surgery and providing significant support on heart surgeries and other medical procedures. He also acknowledges Cathy Jo Land, Haden's wife, for arranging and supporting for the heart and brain surgeries and other medical procedures. Finally, we acknowledge the following people at Springer: Brian Halm, Mary James, Zoe Kennedy, and Donald Menas who were instrumental in the publication of this book.

The title of this book is well chosen: The Art and Science of Machine Intelligence. Like other books on machine learning it explains the science, i.e., the mathematics behind different machine learning techniques. What makes this book unique, however, is the stress on the art. Anyone who has tried to solve real problems using machine learning soon realizes that to be successful one often has to take a hybrid approach, combining various techniques. Many of the examples in the book describe novel forms of ensemble architectures. Furthermore, it highlights how a genetic algorithm designed for subset selection can be used to select the most effective set of features for a machine learning application. The examples are illustrated using difficult real-world problems, including diagnosing Alzheimer's disease based on speech patterns. The final chapter describes a novel method for quantifying the uncertainty area for a trained classifier. The reader will come away from this book with many practical ideas on how to effectively apply machine intelligence to solve difficult problems.

—Dr. Larry J. Eshelman, *IEEE Pioneer in Evolutionary Computation, 2011*

The authors took apart the complex math of Machine Intelligence and weaved it all together in a detailed and understandable way. Their work is an "outstanding" approach to Machine Intelligence that could greatly improve our understanding of the world, especially with medical issues such as Alzheimer's and the better interpretation of complex test results.

—Dr. Michael H. Nance, Corporate Senior Fellow, *Lockheed Martin*

Machine learning from genetic algorithms to very complex networks; new and creative analytic methods are intuitively covered with ample background, examples, and problem sets for the reader.

—Dr. Cihan H. Dagli, Professor, *University of Missouri*

Based on my experience in both the academia and private sectors, this book is an exceptional textbook for a market relevant machine learning graduate level course as well as an ideal technical reference for an experienced data science professional.

—Dr. Brad L. Sims, President, *Capitol Technology University*

Contents

About the Authors

Walker H. Land Jr. After 4 years in the Air Force, receiving technical training and teaching, Walker H. Land, Jr., completed a BS in Physics from Virginia Polytechnic Institute in 1958, then the following 29 years at IBM. While there, he developed technology for the TIROS weather satellite (first ever) and other projects. He received an MS in Engineering and Applied Science from the George Washington University in 1964, completing a thesis on the application of the calculus of variations to the optimization of flight trajectories (recently available online). This led to an assignment at NASA in Huntsville, Alabama, while still working for IBM who had the instrument unit contract for the Saturn—Apollo missions that contained the guidance and control systems, where he was a key person responsible for the translunar flight trajectories for the Apollo missions. He continued at IBM, receiving an award for the most new technology reports submitted by a single individual at the NASA Huntsville headquarters. He retired from IBM in 1990 and began teaching and research at Binghamton University as adjunct lecturer, largely on machine intelligence and complex adaptive systems technologies, many applied to medical problems. Here, he met Professor Schaffer and began the collaboration that led to this book. He retired from Binghamton in 2014 as Emeritus Research Professor, and continues his research and writing.

J. David Schaffer After completing a BS in Aerospace Engineering at the University of Notre Dame in 1964, J. David Schaffer served for 2 years as a member of VISTA, where he gained exposure to problems involving social and economic inequalities. This was followed by an MS in Systems Engineering at Widener University in 1973, where his thesis explored applying graph decomposition methods to city government organization structure. This led to work on analyzing data from clinical trials of psychotropic drugs first at George Washington University, and then at the Tennessee Neuropsychiatric Institute and Vanderbilt University. At Vanderbilt, he completed a PhD in Electrical Engineering in 1984, applying genetic algorithms (GAs) to machine learning, extending GAs into multiobjective applications. The next 25 years he spent with Philips Research extending the theory and

applications of GAs, some to considerable commercial success. In addition to his many published research papers, he holds 43 US patents and serves on the editorial board for the Evolutionary Computation Journal and the steering committee for the bi-annual conference series, Evolutionary Multiobjective Optimization (EMO). In 2010, he retired from Philips and moved to Binghamton University, where he served as Research Professor. Here, he encountered Professor Land and began the fruitful collaboration that led to this book. In 2012, he was designated a Pioneer in Evolutionary Computation by the IEEE Computational Intelligence Society. He is a life member of the IEEE.