SELF-ADAPTIVE SYSTEMS FOR MACHINE INTELLIGENCE

Haibo He

WILEY
A JOHN WILEY & SONS, INC., PUBLICATION
SELF-ADAPTIVE SYSTEMS FOR MACHINE INTELLIGENCE
SELF-ADAPTIVE SYSTEMS FOR MACHINE INTELLIGENCE

Haibo He
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>xi</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>xv</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 The Machine Intelligence Research</td>
<td>1</td>
</tr>
<tr>
<td>1.2 The Two-Fold Objectives: Data-Driven and Biologically Inspired Approaches</td>
<td>4</td>
</tr>
<tr>
<td>1.3 How to Read This Book</td>
<td>8</td>
</tr>
<tr>
<td>1.3.1 Part I: Data-Driven Approaches for Machine Intelligence (Chapters 2, 3, and 4)</td>
<td>8</td>
</tr>
<tr>
<td>1.3.2 Part II: Biologically-Inspired Approaches for Machine Intelligence (Chapters 4, 5, and 6)</td>
<td>9</td>
</tr>
<tr>
<td>1.4 Summary and Further Reading</td>
<td>10</td>
</tr>
<tr>
<td>References</td>
<td>10</td>
</tr>
<tr>
<td><strong>2 Incremental Learning</strong></td>
<td>13</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>13</td>
</tr>
<tr>
<td>2.2 Problem Foundation</td>
<td>13</td>
</tr>
<tr>
<td>2.3 An Adaptive Incremental Learning Framework</td>
<td>14</td>
</tr>
<tr>
<td>2.4 Design of the Mapping Function</td>
<td>19</td>
</tr>
<tr>
<td>2.4.1 Mapping Function Based on Euclidean Distance</td>
<td>19</td>
</tr>
<tr>
<td>2.4.2 Mapping Function Based on Regression Learning Model</td>
<td>20</td>
</tr>
<tr>
<td>2.4.3 Mapping Function Based on Online Value System</td>
<td>23</td>
</tr>
<tr>
<td>2.4.3.1 A Three-Curve Fitting (TCF) Technique</td>
<td>23</td>
</tr>
<tr>
<td>2.4.3.2 System-Level Architecture for Online Value Estimation</td>
<td>26</td>
</tr>
</tbody>
</table>
2.5 Case Study  29
  2.5.1 Incremental Learning from Video Stream  30
    2.5.1.1 Feature Representation  30
    2.5.1.2 Experimental Results  31
    2.5.1.3 Concept Drifting Issue in Incremental Learning  33
  2.5.2 Incremental Learning for Spam E-mail Classification  37
    2.5.2.1 Data Set Characteristic and System Configuration  37
    2.5.2.2 Simulation Results  38
  2.6 Summary  39

References  41

3 Imbalanced Learning  44
  3.1 Introduction  44
  3.2 The Nature of Imbalanced Learning  44
  3.3 Solutions for Imbalanced Learning  48
    3.3.1 Sampling Methods for Imbalanced Learning  49
      3.3.1.1 Random Oversampling and Undersampling  50
      3.3.1.2 Informed Undersampling  51
      3.3.1.3 Synthetic Sampling with Data Generation  52
      3.3.1.4 Adaptive Synthetic Sampling  53
      3.3.1.5 Sampling with Data Cleaning Techniques  56
      3.3.1.6 Cluster-Based Sampling Method  57
      3.3.1.7 Integration of Sampling and Boosting  59
    3.3.2 Cost-Sensitive Methods for Imbalanced Learning  62
      3.3.2.1 Cost-Sensitive Learning Framework  62
      3.3.2.2 Cost-Sensitive Data Space Weighting with Adaptive Boosting  63
      3.3.2.3 Cost-Sensitive Decision Trees  65
      3.3.2.4 Cost-Sensitive Neural Networks  66
    3.3.3 Kernel-Based Methods for Imbalanced Learning  68
      3.3.3.1 Kernel-Based Learning Framework  68
      3.3.3.2 Integration of Kernel Methods with Sampling Methods  69
3.3.3 Kernel Modification Methods for Imbalanced Learning 70
3.3.4 Active Learning Methods for Imbalanced Learning 71
3.3.5 Additional Methods for Imbalanced Learning 73
3.4 Assessment Metrics for Imbalanced Learning 75
  3.4.1 Singular Assessment Metrics 75
  3.4.2 Receiver Operating Characteristics (ROC) Curves 77
  3.4.3 Precision-Recall (PR) Curves 79
  3.4.4 Cost Curves 80
  3.4.5 Assessment Metrics for Multiclass Imbalanced Learning 80
3.5 Opportunities and Challenges 82
3.6 Case Study 84
  3.6.1 Nonlinear Normalization 84
  3.6.2 Data Sets Distribution 88
  3.6.3 Simulation Results and Discussions 92
3.7 Summary 98

References 100

4 Ensemble Learning 108
4.1 Introduction 108
4.2 Hypothesis Diversity 108
  4.2.1 Q-Statistics 109
  4.2.2 Correlation Coefficient 110
  4.2.3 Disagreement Measure 110
  4.2.4 Double-Fault Measure 110
  4.2.5 Entropy Measure 111
  4.2.6 Kohavi–Wolpert Variance 111
  4.2.7 Interrater Agreement 111
  4.2.8 Measure of Difficulty 112
  4.2.9 Generalized Diversity 112
4.3 Developing Multiple Hypotheses 114
  4.3.1 Bootstrap Aggregating 114
  4.3.2 Adaptive Boosting 114
4.3.3 Subspace Learning 119
4.3.4 Stacked Generalization 120
4.3.5 Mixture of Experts 122

4.4 Integrating Multiple Hypotheses 123

4.5 Case Study 126
  4.5.1 Data Sets and Experiment Configuration 127
  4.5.2 Simulation Results 128
  4.5.3 Margin Analysis 129
    4.5.3.1 A Short History of Margin Analysis 129
    4.5.3.2 Margin Analysis for Ensemble Learning 131

4.6 Summary 136

References 137

5 Adaptive Dynamic Programming for Machine Intelligence 140

5.1 Introduction 140
5.2 Fundamental Objectives: Optimization and Prediction 141
5.3 ADP for Machine Intelligence 143
  5.3.1 Hierarchical Architecture in ADP Design 143
  5.3.2 Learning and Adaptation in ADP 146
    5.3.2.1 The Action Network 148
    5.3.2.2 The Reference Network 150
    5.3.2.3 The Critic Network 152
  5.3.3 Learning Strategies: Sequential Learning and Cooperative Learning 154
5.4 Case Study 155
5.5 Summary 160

References 161

6 Associative Learning 165

6.1 Introduction 165
6.2 Associative Learning Mechanism 165
  6.2.1 Structure Individual Processing Elements 166
  6.2.2 Self-Determination of the Function Value 167
  6.2.3 Signal Strength for Associative Learning 168
6.2.4 The Associative Learning Principle 169

6.3 Associative Learning in Hierarchical Neural Networks 173
6.3.1 Network Structure 173
6.3.2 Network Operation 174
   6.3.2.1 Feedforward Operation 174
   6.3.2.2 Feedback Operation 177

6.4 Case Study 180
6.4.1 Hetero-Associative Application 180
6.4.2 Auto-Associative Application 182
   6.4.2.1 Panda Image Recovery 183
   6.4.2.2 Chinese Character Recognition and Recovery 184
   6.4.2.3 Associative Memory for Online Incremental Learning 186

6.5 Summary 187
References 188

7 Sequence Learning 190

7.1 Introduction 190
7.2 Foundations for Sequence Learning 190
7.3 Sequence Learning in Hierarchical Neural Structure 194
7.4 Level 0: A Modified Hebbian Learning Architecture 195
7.5 Level 1 to Level N: Sequence Storage, Prediction, and Retrieval 198
   7.5.1 Sequence Storage 198
   7.5.2 Sequence Prediction 201
      7.5.2.1 Prediction Mechanism 201
      7.5.2.2 Activation of Prediction Neuron 205
      7.5.2.3 Time-Controlled Multiplexer 205
   7.5.3 Sequence Retrieval 207
7.6 Memory Requirement 207
7.7 Learning and Anticipation of Multiple Sequences 208
7.8 Case Study 211
7.9 Summary 212
References 213
The understanding of natural intelligence and developing self-adaptive systems to potentially replicate such a level of intelligence is still one of the greatest unsolved scientific and engineering challenges. With the recent development of brain research and modern technologies, scientists and engineers will hopefully find solutions to develop general-purpose brain-like intelligent systems that are highly robust, adaptive, scalable, and fault tolerant. Yet, there is still a long way to go to achieve this goal. The biggest challenge is how to understand the fundamental principles and develop integrated complex systems to potentially capture those capabilities and eventually bring such a level of intelligence closer to reality.

The goal of this book is to advance the understanding and development of self-adaptive systems for machine intelligence research, and to present the models and architectures that can adaptively learn information, accumulate knowledge over time, and adjust actions to achieve goals. Machine intelligence research draws on theories and concepts from many disciplines, including neuroscience, artificial intelligence, cognitive science, computational theory, statistics, computer science, engineering design, and many others. Because of the inherent cross-disciplinary nature of the research on this topic, most of the materials presented in this book are motivated by the latest research developments across different fields. I hope the research results presented in this book can provide useful and important insights to understand the essential problems of machine intelligence research, and provide new techniques and solutions across a wide range of application domains.

Recent research results have provided strong evidence that brain-like intelligence is very different when compared to traditional artificial intelligence. For instance, although today’s computers can solve very complicated mathematical problems, predict large-scale weather changes, and even win the world chess championship, they use fundamentally different ways of information processing as in the biological brain organism. To this end, this book focuses on the computational foundations and methodologies of machine intelligence toward the “computational thinking” capability for self-adaptive intelligent systems design. Therefore, the research results presented in this book can naturally be grounded as two major parts: data-driven approaches and biologically inspired approaches.

The data-driven approaches aim to understand how to design self-adaptive systems that can autonomously learn from vast amounts of raw data for information and knowledge representation to support the decision-making processes...
within uncertain and unstructured environments, and the biologically inspired approaches target to understand the principles of information processing, association, optimization, and prediction within distributed hierarchical neural network structures. All of these are essential capabilities and characteristics to achieve the general-purpose brain-like machine intelligence in the future. In the last chapter of this book, I also provide a comment about the hardware design for machine intelligence research, which could provide useful suggestions about how to build complex and integrated intelligent systems in massive, parallel, and scalable hardware platforms, such as the dedicated very large scale integration (VLSI) systems and reconfigurable field-programmable gate array (FPGA) technology. Emerging technologies such as memristor are also briefly discussed in the last chapter since such technologies might provide us significant new capabilities to mimic the complexity level of neural structures in the human brain. Furthermore, in order to highlight the wide applications of machine intelligence research, at the end of each chapter, I provide a case study to demonstrate the effectiveness of the proposed method across different domains. These examples should provide useful suggestions about the practical applications of the proposed methods.

This book consists of four major sections, organized as follows:

1. Section 1 (Chapter 1) gives a brief introduction of the self-adaptive systems for machine intelligence research. The research significances and the major differences between traditional computation and brain-like intelligence are presented. A brief review of the book organization and suggested usage is also given in this chapter.

2. Section 2 (Chapters 2, 3, and 4) presents the data-driven approaches for machine intelligence research. The focus is to develop adaptive learning methods to transform a large volume of raw data into knowledge and information representation to support the decision-making processes with uncertainty. Specifically, incremental learning, imbalanced learning, and ensemble learning are presented in this section.

3. Section 3 (Chapters 5, 6, and 7) focuses on biologically inspired machine intelligence research. The goal here is to understand the fundamental principles of neural information processing and develop learning, memory, optimization, and prediction architectures to potentially mimic certain levels of such intelligence. Specifically, adaptive dynamic programming (ADP), associative learning, and sequence learning are discussed in detail.

4. Section 4 (Chapter 8) provides a brief discussion regarding the hardware design for machine intelligence. The goal is to provide some suggestions about the critical design considerations, such as power consumption, design density, and memory and speed requirements, to potentially build such complex and integrated system into real hardware.

This book is intended for researchers and practitioners in academia and industry who are interested in machine intelligence research and adaptive systems development. The presented learning principles, architectures, algorithms, and case
studies will hopefully not only bring the community new insights of machine intelligence research, but it will also provide potential techniques and solutions to bring such a level of capability closer to reality across a wide range of application domains. Furthermore, all the issues discussed in this book are active research topics and present significant challenges to the research community, making this book a valuable resource for graduate students to motivate their own research projects toward their Ph.D. or master-level research. Finally, as machine intelligence research is continuing to attract more and more attention across different disciplines, I also hope this book will provide interesting ideas and suggestions to stimulate undergraduate students and young researchers with a keen interest in science and technology into this exciting and rewarding field; their participation will be critical for the long-term development of a healthy and promising research community.
ACKNOWLEDGMENTS

I am deeply indebted to many colleagues, friends, reviewers, and students who have greatly helped the research development in this field as well the writing of this book.

I am very grateful to my colleagues and friends at the University of Rhode Island (URI) and Stevens Institute of Technology (SIT) for their tremendous support in the development stage of this book. At the URI, many colleagues in the Department of Electrical, Computer, and Biomedical Engineering (ECBE) and the College of Engineering (COE) have provided enormous support in different forms, and I am especially grateful to G. Faye Boudreaux-Bartels, Raymond Wright, Qing (Ken) Yang, Yan (Lindsay) Sun, He (Helen) Huang, Steven M. Kay, Godi Fischer, Leland B. Jackson, Walter G. Besio, Peter F. Swaszek, Frederick J. Vetter, Resit Sendag, Richard J. Vaccaro, Ying Sun, Harish Sunak, Ramdas Kumaresan, Jien-Chung Lo, William J. Ohley, Shmuel Mardix, and Augustus K. Uht for their support for my research and educational development in this area. At Stevens, many friends and colleagues in the Department of Electrical and Computer Engineering (ECE) and Schaefer School of Engineering & Science have provided enormous support for my research development. I am particularly grateful to Joseph Mitola III, Yu-Dong Yao, George Korfiatis, Michael Bruno, Stuart Tewksbury, Victor Lawrence, Yi Guo, Rajarathnam Chandramouli, Koduvayur Subbalakshmi, Harry Heffes, Hong Man, Hongbin Li, Jennifer Chen, Yan Meng, Cristina Comaniciu, and Bruce McNair for their great support for my research development in this field.

I am also deeply indebted to many students and visiting scholars that I have worked with over all these years. Particularly, I would like to thank Sheng Chen, Yuan Cao, Bo Liu, Qiao Cai, Jin Xu, Jie Li, Jian Fu, Jianlong Qiu, Yi Cao, Zhen Ni, Hao Peng, Edwardo A. Garcia, Xiaochen Li, and Yang Bai for their valuable discussions, comments, and proof-reading of the materials presented in this book. Meanwhile, I would also like to thank the students in several classes that I taught regularly, especially the ELE 594: Computational Intelligence and Adaptive Systems and CpE/EE 695: Applied Machine Learning courses, for their helpful suggestions and discussions for the contents related to this book. Although there is no space to mention all their names, this book could not have been written without their help.

Many friends from other universities, research laboratories, and industrial partners have also provided great support for my research development as well as the writing of this book. Particularly, I would like to express my deep gratitude to...
Janusz A. Starzyk for his always great support and help for my research development; many materials presented in this book are inspired by many discussions and joint research efforts. I am also very grateful to Xiaoping Shen for her strong support from mathematical aspect for the machine intelligence research. Furthermore, Venkataraman Swaminathan, Sachi Desai, Shafik Quoraishee, David Grasing, Paul Willson, and many other members at the U. S. Army, Armament Research, Development and Engineering Center (ARDEC), have also provided great support including practical application case studies, real-world data sets, and technical discussions for several research projects over the past several years. I would also like to take this opportunity to thank Charles Clancy, Tim O’Shea, Ray Camisa, and Jeffrey Spinnanger for many stimulating technical discussions at various meetings and their great support for my research development in this field.

I would also like to thank many international experts and scientists who took their precious time to review the materials and provide suggestions for this book. While there is no space to mention all their names, I am particularly grateful to the following experts for their great support: Derong Liu, Jennie Si, Jun Wang, Gary Yen, Robert Kozma, Donald C. Wunsch II, Danil Prokhorov, Marios M. Polycarpou, Mengchu Zhou, Shiejie Cheng, Ping Li, Yaochu Jin, Kang Li, Daniel W Repperger, Wen Yu, Anwar Walid, Tin Kam Ho, Zeng-Guang Hou, Fuchun Sun, Changyin Sun, Robi Polikar, Jinyu Wen, Tiejian Luo, Ying Tan, Xin Xu, Shutao Li, Zhigang Zeng, and many others. Their expertise greatly helped my research development in this area.

I am also very grateful to the support from the U.S. National Science Foundation (NSF) (under grant CAREER ECCS # 1053717), DARPA, and Army ARDEC for their tremendous support for my research development all these years. Their great support has provided me the opportunity to explore all the challenging and exciting research topics in this field.

John Wiley & Sons has provided outstanding support throughout the development stages of this book. I particularly would like to take this opportunity to thank George J. Telecki and Lucy Hitz for their support, valuable suggestions, and encouragement. Without their dedicated help, the writing and production of this book would have taken much longer.

Finally, I would like to extend my deepest gratitude to my family, particularly my wife Yinjiao, for their strong support along the way. I would also like to devote this book to my lovely little one, Eric.

HAIBO HE
CHAPTER 1

Introduction

1.1 THE MACHINE INTELLIGENCE RESEARCH

As the understanding of brain-like intelligence and developing self-adaptive systems to potentially replicate certain levels of natural intelligence remains one of the greatest unsolved scientific and engineering challenges, the brain itself provides strong evidence of learning, memory, prediction, and optimization capabilities within uncertain and unstructured environments to accomplish goals. Although the recent discoveries from neuroscience research have provided many critical insights about the fundamental mechanisms of brain intelligence, and the latest technology developments have enabled the possibility of building complex intelligent systems, there is still no clear picture about how to design truly general-purpose intelligent machines to mimic such a level of intelligence (Werbos, 2004, 2009; Brooks, 1991; Hawkins & Blakeslee, 2004, 2007; Grossberg, 1988; Sutton & Barto, 1998). The challenges of accomplishing this long-term objective arise from many disciplines of science and engineering research, including, but not limited to:

- Understanding the fundamental principles and mechanisms of neural information processing in the biological brain organism.
- Advancement of principled methodologies of learning, memory, prediction, and optimization for general-purpose machine intelligence.
- Development of adaptive models and architectures to transform vast amounts of raw data into knowledge and information representation to support decision-making processes with uncertainty.
- Embodiment of machine intelligence hardware within systems that learn through interaction with the environment for goal-oriented behaviors.
- Design of robust, scalable, and fault-tolerant systems with massively parallel processing hardware for complex, integrated, and networked systems.

To find potential solutions to address all of these challenges, extensive efforts have been devoted to this field from many disciplines, including neuroscience, artificial
introduction, cognitive science, computational theory, statistics, computer science, and engineering design, among others. For instance, artificial neural networks have played an important role in the efforts of modeling functions of brain-like learning (Grossberg, 1988). Backpropagation theory has provided a powerful methodology for building intelligent systems and has demonstrated great success across many domains, including pattern recognition, adaptive control and modeling, and sensitivity analysis, among others (Werbos, 1988a, 1988b, 1990, 2005). There are many other representative works in this field as well, including the memory-prediction theory (Hawkins & Blakeslee, 2004, 2007), reinforcement learning (RL) (Sutton & Barto, 1998), embodied intelligence (Brooks, 1991, 2002), adaptive dynamic programming (ADP) (Werbos, 1997, 1998, 2004, 2009; Si, Barto, Powell, & Wunsch, 2004; Powell, 2007), the “new artificial intelligence” theory (Pfeifer & Scheier, 1999), and others. For instance, recently, a new theoretical framework based on hierarchical memory organization was proposed for designing intelligent machines (Hawkins & Blakeslee, 2004, 2007). This theoretical framework provides potential new solutions for how to understand memory and the prediction mechanism based on the neocortex. Because biological intelligent systems can learn through active interaction with the external environment, reinforcement learning has attracted much attention in the community and demonstrated great success in a wide range of applications (Sutton & Barto, 1998). The key idea of reinforcement learning is to learn how to map situations to actions to maximize the expected reward signal. One of the essential aspects of reinforcement learning is the value function, which specifies “good” from “bad” to guide the goal-oriented behaviors of the intelligent system. For instance, in biological systems, it could be a way of measuring happiness or pain (Starzyk, Liu, & He, 2006). The ideas for embodied intelligence originate from the observation that biological intelligent systems have biological bodies and are situated in a set of realistic environments (Brooks, 1991, 2002). The major research efforts for embodied intelligence are focused on understanding biological intelligent systems, discovering fundamental principles for intelligent behavior, and designing real intelligent systems, including living machines and humanoid robotics. Recently, it is recognized that optimization and prediction play a critical role to bring the brain-like general-purpose intelligence closer to reality (Werbos, 2009). For instance, the recently launched Cognitive Optimization and Prediction (COPN) program from the National Science Foundation (NSF) is a good indication to raise the attention to this critical area by bringing cross-disciplinary teams together to address the essential question of how the brain learns to solve complex optimization and resilient control problems (NSF, 2007). While optimization has a long-standing research foundation in control theory, decision theory, risk analysis, and many other fields, it has specific meanings in terms of machine intelligence research: learning to make better choices to maximize some kind of utility function over time to achieve goals. Extensive research efforts have suggested that ADP is the core methodology, or “the only general-purpose way to learn to approximate the optimal strategy of action in the general case” (Werbos, 2004, 2009). Of course, I would also like to note that many of the aforementioned fields
are strongly connected with each other. For instance, ADP/RL approaches can be “embodied” (e.g., coupled with sensory-motor coordination with active interaction with the external environment) or built in a hierarchical way for effective goal-oriented multistage learning, prediction, and optimization (Werbos, 2009).

From the practical application point of view, recent technology developments have enabled the growth and availability of raw data to occur at an explosive rate, such as sensor networks, security and defense applications, Internet, geographic information systems, transportation systems, weather prediction, biomedical industry, and financial engineering, to name a few. In many of such applications, the challenge is not the lack of the availability of raw data. Instead, information processing is failing to keep pace with the explosive increase of the collected raw data to transform them to a usable form. Therefore, this has created immense opportunities as well as challenges for the machine intelligence community to develop self-adaptive systems to process such vast amounts of raw data for information representation and knowledge accumulation to support the decision-making processes.

To this end, this book focuses on the computational foundations of machine intelligence research toward the “computational thinking” (Wing, 2006) capability for self-adaptive intelligent systems design. For instance, although the traditional artificial intelligence methods have made significant progresses and demonstrated great success across different specific application tasks, many such techniques lack the robustness, scalability, and adaptability across different knowledge domains. On the other hand, biological intelligent systems are able to adaptively learn and accumulate knowledge for goal-oriented behaviors. For instance, although today’s computers can solve very complicated problems, they use fundamentally different ways of information processing than does the human brain (Hawkins & Blakeslee, 2004, 2007; Hedberg, 2007; Sutton & Barto, 1998). That is why a 3-year-old baby can easily watch, listen, learn, and remember various external environment information and adjust his or her behavior, while the most sophisticated computers cannot. In this sense, one may argue that modern computers are just computational machines without intelligence. This raises critical questions such as “What can humans do better than computers, and vice versa?” or, more fundamentally, “What is computable?” from the computational thinking point of view (Wing, 2006). We believe an in-depth understanding of such fundamental problems is critical for machine intelligence research, and ultimately provide practical techniques and solutions to hopefully bring such a level of intelligence closer to reality across different domains.

To give a brief overview of the major differences between traditional computation and brain-like intelligence, Figure 1.1 compares the major characteristics of these two levels of intelligence. One can clearly see that brain-like intelligence is fundamentally different to that of traditional computation in all of these critical tasks. Therefore, from the computational thinking point of view, new understandings, foundations, principles, and methodologies are needed for the development of brain-like intelligence. This book tries to provide the recent advancements in this field to address such critical needs in the community.
INTRODUCTION

1.2 THE TWO-FOLD OBJECTIVES: DATA-DRIVEN AND BIOLOGICALLY INSPIRED APPROACHES

Figure 1.2 illustrates a high-level view of the machine intelligence framework that we focus on in this book. Here, there are two important components: the intelligent core such as neural network organizations and learning principles, and the interaction between the intelligent core and the external environment through sensorimotor pathways (embodiment). To this end, this book includes
two major parts to address the two-fold objectives: data-driven approaches and biologically inspired approaches for machine intelligence research. This will not only allow us to understand the foundations and principles of the neural network organizations and learning within the intelligent core, but it also allows us to advance the principled methodologies with a focus on the data processing path (sensing, acquisition, processing, and action). The key is to understand how a brain-like system can adaptively interact with unstructured and uncertain environments to process vast amounts of raw data to develop its internal structures, build associations and predictions, accumulate knowledge over time, and utilize self-control to achieve goals.

The underlying motivation of data-driven approaches is quite straightforward: Data provide the original sources for any kind of information processing, knowledge transformation, and decision-making processes. From the computational intelligence point of view, data are almost involved in every aspect of “intelligence”: reasoning, planning, and thinking, among others. Therefore, data can be a vital role for machine intelligence development in different formats, such as sensing, acquisition, processing, transformation, and utilization. You can think about many examples in real-world applications from this perspective, ranging from picking up a pen from your office desk, to driving a car in the metropolitan area of New York City, to scheduling your calendar for the next month. All of these tasks involve data analysis at different levels. If one would like to design an intelligent machine to possibly replicate certain levels of brain-like intelligence, many critical questions are raised from the data computational point of view, such as: What kind of data are necessary to support the decision-making processes? How can an intelligent machine continuously learn from non-stationary and noisy data? How do you effectively combine multiple votes from different hypotheses based on different data spaces for optimal decisions?

Specifically, in this book we will discuss the following data-driven approaches for machine intelligence research:

- Incremental Learning. Incremental learning is critical to understand brain-like intelligence and potentially bringing such a level of intelligence closer to reality in at least two aspects. First, intelligent systems should be able to learn information incrementally throughout their lifetimes, accumulate experience, and use such knowledge to benefit future learning and decision-making processes. Second, the raw data that come from the environment with which the intelligent system interacts becomes incrementally available over an indefinitely long (possibly infinite) learning lifetime. Therefore, the learning process in such scenarios is fundamentally different from that of traditional static learning tasks, where a representative data distribution is available during the training time to develop the decision boundaries used to predict future unseen data. Furthermore, how to achieve global generalization through incremental learning is a crucial component in the correct understanding of such problems. Therefore, it is critical to go beyond the conventional “compute–store–retrieve” paradigm for the development of
natural intelligent systems for such large-scale and complicated data processing systems.

• Imbalanced Learning. In many real-world applications, an intelligent system needs to learn from skewed data distributions to support decision-making processes. Such skewed distribution with underrepresented data may significantly compromise learning capability and performance. For instance, many of the existing learning algorithms assume or expect balanced data distributions to develop the decision boundary. Therefore, when presented with the imbalanced data, such learning algorithms fail to properly represent the distributive characteristics of the data and resultantly provide worse learning performance. Due to the inherent complex characteristics of imbalanced data and its wide occurrence in many real systems, the imbalanced learning problem has presented a significant new challenge to society with wide-ranging and far-reaching application domains.

• Ensemble Learning. Generally speaking, ensemble learning approaches have the advantage of improved accuracy and robustness compared to the single model–based learning methods. In the ensemble learning scenario, multiple hypotheses are developed and their decisions are combined by a voting method for prediction. Since different hypotheses can provide different views of the target function, the combined decision will hopefully provide more robust and accurate decisions compared to single model–based learning methods. There are two critical aspects related to ensemble learning. First, how can one develop multiple hypotheses given the training data? For instance, to obtain the diversified hypotheses, many techniques such as bootstrap aggregating (bagging), adaptive boosting, random subspace, stacked generalization, mixture of experts, and others have been proposed in the community. Second, how can one effectively integrate multiple hypotheses outputs for improved final decisions? This mainly includes different types of combinational voting strategies, which will also be discussed in this book.

In addition to the data-driven approaches, we have a keen interest to understand and develop biologically inspired approaches for machine intelligence. Recent brain science research has provided strong evidence that the biological brain uses fundamentally different ways in handling various tasks than today’s computers (Hawkins & Blakeslee, 2004, 2007; Hedberg, 2007). For instance, although IBM’s Deep-Blue can win a chess game against a world champion over a decade ago, it did not tell us too much about the development of general-purpose brain-like intelligent machines as it uses completely different ways of information processing as in the human brain. On the other hand, the evolutionary algorithm has recently showed great potential to develop the self-learning capabilities for a master-level chess program, which could provide us important insights to understanding the essence of machine intelligence (Fogel, Hays, Han, & Quon, 2004). From this perspective, the important question is how to develop biologically inspired system-level models and architectures that are able
to mimic certain levels of brain intelligence. In this book we will discuss three major components on this.

- Adaptive Dynamic Programming (ADP). ADP has been widely recognized as the key methodology to understand and replicate general-purpose brain-like intelligence in the community (Werbos, 1994, 1997, 2004, 2009; Si et al., 2004; Powell, 2007). There are two major goals of the ADP research that can contribute to the machine intelligence research: optimization and prediction. Specifically, optimization in this case can be defined as learning to make better choices to maximize some kind of utility function over time to achieve goals (Werbos, 2009). To this end, the foundation for optimization over time in stochastic processes is the Bellman equation (Bellman, 1957), closely tied with the Cardinal utility function concept by Von Neumann. In addition to optimization, recently strong evidence from neurobiology research suggested that prediction is another equally important component to provide a level of adaptive general-purpose intelligence (Werbos, 2009). Prediction in the ADP design can be considered in a more general way to include much important information, such as the future sensory inputs from observed data as well as modeling and reconstructing of the unobserved state variable, with the objective of facilitating action selection toward optimization (Werbos, 2009). In this book, we propose a hierarchical ADP architecture with multiple-goal representations to effectively integrate optimization and prediction together for machine intelligence research.

- Associative Learning. Associative memory plays a critical role for natural intelligence based on information association and anticipation. Generally speaking, there are two types of associative memories: hetero-associative and auto-associative memory. Hetero-associative memory makes associations between paired patterns, such as words and pictures, while auto-associative memory associates a pattern with itself, recalling stored patterns from fractional parts of the pattern. It is believed that the human brain employs both hetero-associative and auto-associative memory for learning, action planning, and anticipation (Rizzuto & Kahana, 2001; Brown, Dalloz, & Hulme, 1995; Murdock, 1997). The memory evolved in the human brain is self-organized, hierarchically distributed, and data driven. For instance, self-organization is responsible for formation of hierarchically organized structures not only in the human brain but also in the nervous systems of lower vertebrates (Malsburg, 1995). In this book, I will focus on the essential characteristics of associative learning, including self-organization, sparse and local connection, and hierarchical structure.

- Sequence Learning. Sequence learning is widely considered among one of the most important components of human intelligence, as most human behaviors are in the sequential format, including, but not limited to, natural language processing, speech recognition, reasoning and planning, and others. Therefore, the understanding of fundamental problems of sequence learning could provide us critical insights for machine intelligence research.