Lecture Notes in Computer Science 1387

Adaptive Processing of Sequences and Data Structures

Bearbeitet von
C.Lee Giles, Marco Gori

1. Auflage 1998. Taschenbuch. xiv, 438 S. Paperback
ISBN 978 3 540 64341 8
Format (B x L): 15,5 x 23,5 cm
Gewicht: 1390 g

Foreword

This book was originated from the International Summer School on “Adaptive Processing of Sequences and Data Structures” held in Vietri sul Mare, Salerno (Italy) from September 6 to 13. Fourteen lecturers and about sixty students, mainly post-docs, interacted in a very lively and exciting environment. In addition to the lectures there was a nice panel session on applications and a poster session reserved for students. Because of the broad range of topics covered in the school, the lectures were mainly tutorial, and raised an impressive number of questions that have certainly contributed to improve the quality of the presentation of this book.

This is the second initiative jointly instituted by the International Institute for Advanced Studies “Eduard Renato Caianiello” (IIASS) and the “Ettore Majorana” Foundation and Centre for Scientific Culture (EMFCSC) in the memory of E.R. Caianiello, a pioneer in the field of neural networks and active researcher in many related domains. The first initiative of this kind was held in Erice from September 27 to October 7, 1996 on “Learning in Graphical Models” and was directed by M.I Jordan and D. Heckerman.

The school has been organized thanks to the support of many sponsors: the IIASS, the University of Salerno, the AI*IA (Italian Association for Artificial Intelligence), the SIREN (Italian Neural Network Society), the Italian chapter of the IEEE Neural Network Council, and the Fondazione Bordoni (Rome). I thank them all and I'm delighted to share this book with you.

M. Marinaro, IIASS and University of Salerno

Preface

Time is commonly perceived by humans as a relentless drummer which beats out the rhythm of life. While philosophers have always been interested in capturing the essence of time, most scientists have simply regarded time as a fundamental variable interwound with physical and biological laws, as well as with information processing. Human beings apparently rely on a somewhat sequential view of time not only for everyday life but also for performing mathematical operations and, more generally, symbolic information processing. Neurobiologists and psychologists, however, have clear evidence that human information processing takes place in the brain by means of neuronal units, strongly interconnected by synaptical connections. For many years, computer architects have designed computers mainly relying upon the sequential view of time and, consequently of information processing. This perspective, however, might change in the next few years. When discussing parallel processing, Patterson and Hennessey in chapter 9 of their classic textbook *Computer Organization & Design. The Hardware/Software Interface*, quote a very nice précis, which is in fact an Irish proverb: "There are finer fish in the sea than have ever been caught." Unfortunately, this fish cannot be easily caught. Most classic concepts from computer science have a inherently sequential nature which seems to be inherited from the notion of algorithm. On the other hand, in the connectionist community, the most established models of learning are based on a truly parallel data processing scheme that, however, is assumed to deal with static data types. Hence, time plays a crucial role in the development of most effective models of both parallel processing and learning. Most interesting intelligent information processing tasks, like speech recognition, are based on the presence of some form of structure that, however, cannot easily be grasped at a symbolic level. The way humans process this information can be regarded neither as strictly symbolic nor subsymbolic, neither sequential nor parallel, but is in fact a sort of magic mixture of all of these. Not only is the human brain a complex graph of elementary units, but data to be processed can often be regarded as complex structures of elementary units themselves. In these cases, a purely sequential view of time is not necessarily appropriate to account for the parallel computation that can take place in different pieces of complex data structures. Finally, some interesting temporal processes can be more conveniently modeled in the deterministic framework and others are better modeled in the stochastic framework.

In the last few years, this very wide perspective has stimulated many researchers with different backgrounds to find appropriate architectures and effective learning procedures. In particular, in the field of neural networks, this interest is witnessed by workshops, like those that have taken place in the last few years at the Neural Information Processing System conference and by special issues of technical journals (IEEE Trans. on Neural Networks, March 1994, Giles, Kuhn, and Williams (eds.); ACM SIGART Bulletin, July 1994, Chappelier and Grumbach (eds.); Neurocomputing, July 1997, Gori, Mozer, Tsoi, and Watrous (eds.)).

It was in the context of the above remarks that this book originated. The book presentss in fact the invited lectures given at the Caianiello Summer School on Adaptive Processing of Sequences and Data Structures, which took place in a beautiful setting in Vietri sul Mare (Salerno), Italy, 6–13 September, 1997; we are now delighted to share it with you. It is organized in fifteen chapters grouped into five sections: architectures and learning in recurrent networks, adaptive processing of data structures, probabilistic models, analog versus digital computation, and applications.

The theme of the three chapters in the first group evolves around the foundations of architectures and learning in recurrent neural networks. To some extent, these chapters extend naturally the basic concepts of multilayer perceptrons. However, it is clearly stated that recurrent networks are more complex in terms of architecture and that they can give rise to a number of qualitatively different behavior patterns. Moreover, the need to learn temporal dependencies gives rise to additional difficulties that become very serious when using classic gradient-based learning algorithms. The first two chapters in this group, by A.C. Tsoi, provide a systematic treatment of different architectures, and give the foundations of gradient-based learning methods. Different architectures are related depending on their computational power and, likewise, many learning algorithms are reviewed in a unified framework. The third chapter, by Wan and Beaufays, presents an approach to deriving gradient algorithms for time-dependent neural networks that is based on a set of simple block diagram manipulation rules.

The second group of three chapters is about adaptive processing of data structures. As already pointed out, in many real-word problems, data are not only significantly structured, but many composing features have a subsymbolic nature. A structured organization of information is typically required by symbolic processing. On the other hand, most connectionist models assume that data are organized according to relatively poor structures, like arrays or sequences. The first chapter of this group, by P. Frasconi, proposes a general framework for extending the processing of sequences to directed acyclic graphs using the concept of a *recursive network*, and gives a brief description of models for learning recursive transitions, based on either neural networks or probabilistic graphical models. The second chapter, by A. Sperduti, focuses attention on neural models for processing data structures. Most theoretical results are reviewed on approximation capabilities, circuit complexity, and learning. The third chapter, by M. Gori, gives hints on the complexity of loading the weights in simple static and recursive networks, for processing either sequences or data structures.

The third section of the book deals with probabilistic models of temporal processes. The underlying assumption is that in most realistic scenarios, the sequences can hardly be modeled in a deterministic setting, like the one assumed in the first chapters. In many application domains, ranging from time series prediction to speech recognition, researchers have been debating which setting (deterministic or probabilistic) is more appropriate to deal with specific learning problems. In spite of the discussions we have seen in the scientific community (e.g., HMMs versus recurrent networks for speech recognition), apart from specific tasks, it does not seem easy to come up with a definite answer. The first chapter of this group, by Z. Ghaharamani, presents a probabilistic framework for learning models of temporal data, which is based on Bayesian networks. Classic models like HMMs are seen in a more general perspective and, moreover, different approaches are given to handle the intractability of the inference in dynamic Bayesian networks. The second chapter, by P. Baldi, looks at time series associated with neuronal signals, provides a brief overview of some of the main problems and applications in this area, and develops a general probabilistic approach for the study of spike trains. The third chapter, by L. Parra, discusses various optimization principles based on statistical independence focusing in particular on some aspects of time.

The fourth section of the book explores connections between analog and digital computation. When discussing the links between the human brain and computers, J. von Neumann pointed out that "... processes which go through the nervous system may change their character from digital to analog, and back to digital, etc., repeatedly." Interestingly enough, in today's artificial neural networks, and especially in recurrent neural networks, we can clearly perceive the dual character mentioned by von Neumann: the truly analog character of the model can be given a nice digital interpretation through appropriate approximations of the nonlinear dynamics. This is shown in the first two chapters that are devoted to exploring the links between recurrent networks and automata. In particular, the first chapter, by M. Maggini, is a survey on the research carried out especially in the last five years concerning the extraction of automata from recurrent networks. Special emphasis is placed on finite state automata and algorithms for inductive inference are also given under noisy examples. The second chapter, by Sun, Giles, Chen, and Lee, goes beyond the links with finite automata by introducing the *NNPDA neural network pushdown automata*. It is shown that the *NNPDA* can successfully infer from examples context-free grammars related to simple languages (e.g., the balanced parenthesis language).

The last section of the book is more focused on applications. In the first chapter, by M. Mozer, two different ways of parsing the stream of time are proposed, namely the *clock-based segmentation* and the *event-based segmentation*. It is pointed out that a complex control problem that appears intractable when cast in terms of a clock-based segmentation, has instead a straightforward solution when cast in terms of an event-based segmentation. A real-world application is shown concerning an adaptive control of indoor lighting in a residence. In the second chapter, by Bourlard and Morgan, an overview is given and new research directions on hybrid HMM/ANN systems for speech recognition are discussed. The chapter shows a well-known integration of HMMs, used for capturing the temporal structure of speech, with multilayer perceptrons, used to predict phones from acoustic vectors. New theoretical developments in training multilayer perceptrons to maximize the posterior probabilities of the correct models for speech utterances are also given. The last chapter, by P. Gallinari, presents predictive models conceived for capturing some kind of dependency inside non-stationary sequences. Application to speech and character recognition are presented which achieve the state of the art performance.

This book could not have been completed without the full-hearted cooperation of all the authors. The lively discussion during the Summer School among lecturers and students provided stimulating feedback for improving the drafts distributed to the students; we are thankful to them all. We are also grateful to Professor M. Marinaro who was mainly in charge of the organization of the Summer School and arranged for the IIASS (International Institute for Advanced Scientific Studies) research center to be the venue of the event.

January 1998 C. Lee Giles and Marco Gori