SYMBOLIC AND CONNECTIONIST LEARNING TECHNIQUES FOR GRAMMATICAL INFERENCE

Autor: René Alquézar Mancho
Director: Alberto Sanfeliu Cortés

March 1997
A Ana, a Marc y Jéssica, y a mis padres
Acknowledgements

This doctoral thesis has been developed in the framework of a predoctoral research grant given by the Generalitat de Catalunya (by Resolution dated 29/10/91 and published in DOGC-1514-6/11/91) and a CICYT project (CICYT TAP96-0629-C04-02), that have provided economical and infrastructure support for this work. This support is gratefully acknowledged.

I would like to express my thankfulness to all the people who have contributed in one way or another to the work presented in this doctoral thesis. In particular, I would like to mention:

Dr. Alberto Sanfeliu, Professor in the Department of Enginyeria de Sistemes, Automàtica i Informàtica Industrial at the Universitat Politècnica de Catalunya (UPC) and researcher in the Institut de Robòtica i Informàtica Industrial (CSIC-UPC), for his continuous advice and guidance along the realization of this dissertation. He gave me his support in every single moment, something that I will never forget, and he also showed me to enjoy scientific research. My sincere acknowledgements and admiration to him.

Dr. Rafael Huber and Dr. Pere Caminal, Heads of the Institut de Robòtica i Informàtica Industrial and the Institut de Cibernètica, respectively, for their kind support during the evolution of this dissertation. They provided me with all the resources I needed of these research institutions.

Dr. Pere Brunet and Dr. Fernando Orejas, who have been successively Head of the Department of Llenguatges i Sistemes Informàtics (LSI) at the UPC since the time I entered this department, as well as Dr. Núria Castell, Head of the Artificial Intelligence Section, for giving me the necessary support to carry out this thesis. Likewise, I would like to thank the Department of LSI as a whole for offering me some teaching load reductions that have enabled me to devote more time to this work.

Dr. Ton Sales, Professor in the Department of LSI, who has been my tutor during the development of this dissertation, for his guidance in doctoral courses and his enjoyable lectures I have had the chance to attend.
Acknowledgements

Dr. Josep Maria Sopena, Professor in the Department of Psicologia Bàsica at the Universitat de Barcelona, for helping me to introduce into the field of neural networks, as well as for their valuable comments and opinions during the numerous meetings that we held together.

Dr. Paco Casacuberta and Dr. Enrique Vidal, Professors at the Universitat Politècnica de València, and also Dr. José Oncina and Dr. Mikel Forcada, Professors at the Universitat d'Alacant, for their estimable comments and scientific aid.

Dr. Pierre Dupont, currently in the Department of Computer Science at Carnegie Mellon University, for kindly providing me with his benchmark data for my experiments.

María Asunción Castaño (I hope to call her Dr. soon), Professor in the Department of Informàtica at the Universitat Jaume I, for her friendship and for sharing with me a common interest and activity in the areas of grammatical inference and neural networks.

Miguel Sainz, Jordi Cueva, Martí Font, and Daniel Casillas, all of them students and members of the research group led by Dr. Sanfeliu, for their collaboration in multiple aspects of the work and for their good mood and solidarity.

My sincere gratitude goes to my friend and office mate Angela Nebot, who is also Dr. and Professor in the Department of LSI, for her support and for sharing with me a lot of hard and happy moments during these years. I wish her all the best, she truly deserves it.

I would also like to thank the members (including professors, research fellows, administrative and technical staff) of the Institut de Cibernètica, the Institut de Robòtica i Informàtica Industrial, and the Department of LSI, who have been my companions during the latest years. Their support and friendship is highly appreciated.

Finally, I want to mention all the people that provided moral support in every single moment, especially my wife Ana, my parents René and Aurea, and my mother-in-law Maria. They gave me courage to continue with this research during the hard moments that all too often exist during the development of a doctoral thesis. I do not forget my children Marc and Jessica, whom I must ask their forgiveness for all the time I could not be with them due to this work.
Summary of the doctoral thesis

During the last thirty years, learning has become a foremost topic within the field of artificial intelligence and a great deal of effort has been devoted to it by researchers. The topic of learning has been divided into several areas as progress in artificial intelligence has developed. One of them is related to inductive inference, which can be defined as "the process of inferring general rules from examples, or theories from facts, carried out by an agent or system".

A great part of the work on the theory and methods for inductive inference has been done within Gold's general paradigm of inductive inference, in which an agent tries to identify a concept from a determined class of concepts or object space, subject to a certain success criterion, by collecting data and guessing a description of the concept, which is selected from a hypothesis space. Normally, the input data consist of examples (and maybe also counterexamples) of the target concept. The term grammatical inference (GI) has been used in this context to denote the inductive inference of classes of languages described by some appropriate representations like grammars or automata.

Although some theoretical limitations of grammatical inference are known, several practical GI methods have been proposed in the literature for a variety of language classes, typically regular languages or some subclass of regular languages, using both the classical symbolic paradigm and the more recent connectionist approach, in which some type of artificial neural network is used. Since GI can provide automatic methods for learning models of pattern classes in syntactic pattern recognition tasks, the interest in the development of powerful and flexible GI tools is clear. Indeed, some applications of GI have been reported in the areas of natural language modelling and translation, speech recognition, and computer vision.

The aim of the work developed in this dissertation was to study both symbolic and connectionist learning techniques for grammatical inference and their relationships, trying to develop new GI methods and/or improvements on the existing ones. The research on GI methods was restricted to the scope of two classes of languages, the regular languages and (some subclass of) the context-sensitive languages (CSLs), and to the case of learning from examples (positive and negative samples).
Summary of the doctoral thesis

This thesis is structured in four parts for a total of ten chapters. The first part, *introduction and review* (Chapters 1 to 4), presents an extensive state-of-the-art review of both symbolic and connectionist GI methods, that serves also to state most of the basic material needed to describe later the contributions of the thesis. These contributions constitute the contents of the rest of parts (Chapters 5 to 10).

An introduction to the topic and objectives of the work is given in Chapter 1. The known theory and (symbolic) methods for regular GI are described in Chapter 2, while the symbolic methods proposed for the inference of context-free languages (CFLs), CSLs, and pattern languages are reviewed in Chapter 3. Chapter 4 introduces the neural network architectures, learning algorithms, and training procedures that have been used in the connectionist works on GI reported previously.

The second part, *contributions on symbolic and connectionist techniques for regular grammatical inference* (Chapters 5 to 7), describes the contributions related to the theory and methods for regular GI, which include other lateral subjects such as the representation of finite-state machines (FSMs) in recurrent neural networks (RNNs).

In Chapter 5, the so-called unbiased finite-state automata (or UFSAs) are defined and proposed as hypotheses for the problem of learning regular languages from both positive and negative examples. The theory of regular GI is reformulated according to this representation, and both non-incremental and incremental methods using UFSAs are presented and evaluated on a benchmark test. It must be noted that until very recently, there has been a lack of incremental methods for regular GI from both positive and negative examples. Moreover, in contrast with other methods, the consistency analysis provided by the use of UFSAs is symmetric and might easily be extended to the problem of learning recognizers of several non-overlapping classes (represented by regular languages) from examples belonging to these classes.

In Chapter 6, two (heuristic) connectionist approaches to regular GI based on RNNs are discussed, which involve training RNNs to learn the next-symbol prediction task (from a stochastic presentation of positive examples) or the string classification task (from a given set of positive and negative examples), respectively, and extracting afterwards an UFSA from the trained net. These approaches have been also validated on some benchmark tests used by different researchers, and the effects of varying the network architecture and activation functions have been assessed.

The proposed UFSA extraction methods improve other methods reported previously to extract an automaton from a trained RNN in the sense that they guarantee the obtaining of a consistent deterministic automaton in a single run. Another nice property is that they use an inductive scheme which is similar to the one used by most of the symbolic methods for regular GI (i.e. state merging...
from a canonical automaton). On the other hand, it is shown that by using some activation functions in RNNs different from the commonly employed sigmoid function the learning performance of these networks can be improved considerably, and a theoretical justification of this empirical result is given.

In Chapter 7, an algebraic framework to represent FSMs in RNNs is presented that unifies and generalizes some of the previous proposals and serves to explain the different representational capabilities of first-order and higher-order single-layer RNNs. This framework can be used to insert symbolic knowledge in RNNs prior to learning from examples and to keep this knowledge while training the network, and it has given rise to a hybrid semi-automated GI methodology, called active grammatical inference (AGI), which is described. In contrast with previous approaches, the insertion method proposed is valid for a wide range of activation functions, whenever some stability conditions are met.

The third part of the thesis, augmented regular expressions and their inductive inference, comprises Chapters 8 and 9. The augmented regular expressions (or AREs) are defined and proposed as a new representation for a subclass of CSLs that does not contain all the context-free languages but a large class of languages capable of describing patterns with symmetries and other (context-sensitive) structures of interest in pattern recognition problems.

In Chapter 8, AREs and their components are formally defined, and an efficient method to recognize a string as belonging or not to the language represented by an unambiguous ARE is described in detail. It is also demonstrated that AREs cover the class of pattern languages defined by Angluin. AREs constitute a better representation than context-sensitive grammars (CSGs) for the associated subclass of CSLs, since the language description is more intelligible, the recognition method is more efficient (at least for unambiguous AREs), and to the contrary of CSGs, practical learning algorithms to infer AREs from examples can be devised.

In Chapter 9, the inductive inference of AREs from string examples is discussed. A general method to infer AREs from examples is proposed that is based on a regular GI step followed by a constraint induction process. In addition, a specific method for learning AREs from positive examples, in which a connectionist regular GI technique is used, is presented, and the results of the application of this method to the inference of a set of eight test CSLs are reported.

It must be emphasized that work on CSL learning has been extremely scarce in the literature of GI and there is a lack of methods to infer descriptions of CSLs. However, there is a need for GI methods capable of inferring CSLs, specially for pattern recognition tasks in computer vision, where objects usually contain structural relationships that are not describable by regular languages or CFLs.
The fourth part of the thesis just includes Chapter 10: conclusions and future research. Chapter 10 summarizes the main results obtained and points out the lines of further research that should be followed both to deepen in some of the theoretical aspects raised and to facilitate the application of the developed GI tools to real-world problems in the area of computer vision.

The research effort documented in this dissertation has focused on methodological and theoretical issues. Therefore, the different empirical studies presented in this doctoral thesis serve only to demonstrate the feasibility and validity of the proposed methods. Nevertheless, prototype systems that use the GI techniques presented here have already been reported for some computer vision applications (e.g. recognition of traffic signs in outdoor scenes) by other members of our research group.
Index

1 Introduction ............................................. 1

1.1 Learning and inductive inference ......................... 2

1.2 Success criteria for inductive inference ...................... 4

1.3 Restrictions on inductive inference methods and some general methods. 7

1.4 Grammatical inference ..................................... 12

1.5 Neural networks and their use for grammatical inference ............. 15

1.6 Applications of grammatical inference ....................... 18

1.6.1 Natural language modelling and translation ................. 19

1.6.2 Speech recognition ......................................... 20

1.6.3 Computer vision ........................................... 23

1.7 Objectives of the thesis ..................................... 25

1.8 Structure of the thesis ....................................... 26

2 Regular grammatical inference: theory and symbolic methods 31

2.1 Basic concepts, definitions and notation ...................... 32
2.1.1 Languages and grammars ........................................ 32
2.1.2 Finite-state machines and finite-state automata ............ 34
2.1.3 Regular expressions ............................................ 36

2.2 Basic theory for regular grammatical inference .................. 38
  2.2.1 Statement of the RGI problem ................................ 39
  2.2.2 Further definitions ........................................... 39
    2.2.2.1 Derived automata ....................................... 40
    2.2.2.2 Canonical automaton of a regular language, $k$-tails and DFA minimization .............. 40
    2.2.2.3 Structural completeness of a sample .................... 41
    2.2.2.4 Automata related to a positive sample .................. 42
    2.2.2.5 Lattices of automata .................................... 43
  2.2.3 The search space of the RGI problem ......................... 44
  2.2.4 The complexity of regular language identification .......... 46

2.3 Symbolic methods for regular grammatical inference ............ 48
  2.3.1 Methods for RGI from a positive sample ..................... 48
    2.3.1.1 Heuristic techniques ................................... 49
    2.3.1.2 Characterizable techniques ............................ 53
  2.3.2 Methods for RGI from a positive and a negative sample .... 55
  2.3.3 Methods for RGI using queries ................................ 58
3 Non-regular grammatical inference through symbolic approaches 63
  3.1 Context-free grammatical inference (CFGI) 64
    3.1.1 CFGI from positive structural examples 66
    3.1.2 CFGI from unstructured data 73
      3.1.2.1 Inference of even-linear grammars 73
      3.1.2.2 Inference of other subclasses of CFGs from positive examples 75
      3.1.2.3 Inference of CFGs from positive and negative examples 77
    3.1.3 Inference of stochastic CFGs 82
    3.1.4 Inference of CFGs from queries 85
  3.2 Inference of controlled grammars 87
  3.3 Inference of transition networks 90
    3.3.1 Inference of BTN\textsc{s} 90
    3.3.2 Inference of ATNs 93
  3.4 Inference of pattern languages 96
    3.4.1 Inference of pattern languages from positive examples 97
    3.4.2 Inference of pattern languages using queries 101
  4 Grammatical inference through connectionist approaches 107
    4.1 Architectures and learning algorithms 108
      4.1.1 Single-layer recurrent neural networks (SLRNNs) 109
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.1.1</td>
<td>First-order SLRNNs</td>
<td>110</td>
</tr>
<tr>
<td>4.1.1.2</td>
<td>Second-order SLRNNs</td>
<td>119</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Augmented single-layer recurrent neural networks (ASLRNNs)</td>
<td>125</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Recurrent Cascade-Correlation (RCC) architecture</td>
<td>130</td>
</tr>
<tr>
<td>4.1.4</td>
<td>The DOLCE architecture</td>
<td>132</td>
</tr>
<tr>
<td>4.1.5</td>
<td>First-order 2-layer RNNs</td>
<td>137</td>
</tr>
<tr>
<td>4.2</td>
<td>Grammatical inference and recognition using recurrent neural networks</td>
<td>138</td>
</tr>
<tr>
<td>4.2.1</td>
<td>GI from positive examples: the next-symbol prediction task</td>
<td>138</td>
</tr>
<tr>
<td>4.2.2</td>
<td>GI from positive and negative examples: the string classification task</td>
<td>145</td>
</tr>
<tr>
<td>4.3</td>
<td>FSA extraction from RNNs for regular GI</td>
<td>154</td>
</tr>
<tr>
<td>4.4</td>
<td>Connectionist approaches to context-free GI</td>
<td>160</td>
</tr>
<tr>
<td>4.4.1</td>
<td>The neural network pushdown automaton (NNPDA)</td>
<td>162</td>
</tr>
<tr>
<td>4.4.1.1</td>
<td>NNPDA architecture and learning algorithm</td>
<td>164</td>
</tr>
<tr>
<td>4.4.1.2</td>
<td>Context-free grammatical inference through NNPDA</td>
<td>169</td>
</tr>
<tr>
<td>4.4.1.3</td>
<td>Extraction of pushdown automata from trained NNPDA</td>
<td>171</td>
</tr>
<tr>
<td>4.4.1.4</td>
<td>Some experimental CFGI results using NNPDA</td>
<td>172</td>
</tr>
<tr>
<td>4.4.2</td>
<td>A connectionist symbol manipulator that induces CFG production rules</td>
<td>176</td>
</tr>
<tr>
<td>4.5</td>
<td>Concluding remarks</td>
<td>181</td>
</tr>
<tr>
<td>5</td>
<td>Regular grammatical inference from positive and negative data using</td>
<td>183</td>
</tr>
<tr>
<td></td>
<td>unbiased finite state automata</td>
<td></td>
</tr>
<tr>
<td>Index</td>
<td>Chapter</td>
<td>Title</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>5.1</td>
<td>Unbiased Finite State Automata</td>
<td>184</td>
</tr>
<tr>
<td>5.2</td>
<td>Basic theory for regular grammatical inference using UFSAs</td>
<td>189</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Further definitions and theory about UFSAs</td>
<td>189</td>
</tr>
<tr>
<td>5.2.1.1</td>
<td>Canonical UFSA of a regular language and universal UFSAs</td>
<td>189</td>
</tr>
<tr>
<td>5.2.1.2</td>
<td>UFSAs related to a sample</td>
<td>189</td>
</tr>
<tr>
<td>5.2.1.3</td>
<td>The extends and consistently-extends predicates</td>
<td>190</td>
</tr>
<tr>
<td>5.2.1.4</td>
<td>Derived UFSAs and lattices of UFSAs</td>
<td>191</td>
</tr>
<tr>
<td>5.2.1.5</td>
<td>Structural completeness of a sample with respect to an UFSA</td>
<td>192</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Restatement of the RGI problem</td>
<td>193</td>
</tr>
<tr>
<td>5.2.3</td>
<td>The search space of the RGI problem using UFSAs</td>
<td>194</td>
</tr>
<tr>
<td>5.3</td>
<td>Basic method for non-incremental RGI from positive and negative examples</td>
<td>197</td>
</tr>
<tr>
<td>5.4</td>
<td>Pseudo-incremental methods for RGI using UFSAs</td>
<td>202</td>
</tr>
<tr>
<td>5.5</td>
<td>Experimental assessment</td>
<td>208</td>
</tr>
<tr>
<td>5.5.1</td>
<td>RGI from sparse samples</td>
<td>208</td>
</tr>
<tr>
<td>5.5.2</td>
<td>RGI from complete samples</td>
<td>212</td>
</tr>
<tr>
<td>6</td>
<td>Regular grammatical inference using recurrent neural networks</td>
<td>215</td>
</tr>
<tr>
<td>6.1</td>
<td>UFSA extraction from trained RNNs for regular grammatical inference</td>
<td>217</td>
</tr>
<tr>
<td>6.2</td>
<td>Effect of different activation functions on learning performance of RNNs</td>
<td>226</td>
</tr>
</tbody>
</table>
6.2.1 Alternatives to the use of the sigmoid function as activation function in RNNs .......................... 227

6.2.2 An empirical study on the next-symbol prediction task .................. 240
  6.2.2.1 Architectures and learning algorithms used in the tests 240
  6.2.2.2 Procedure and results .................................. 240

6.3 RGI from positive and negative examples by training RNNs to learn the string classification task: experimental assessment ............. 246
  6.3.1 Architectures and algorithms used in the tests ....................... 246
  6.3.2 Procedure and results .................................. 248

7 Representation of finite-state machines in recurrent neural networks and the active grammatical inference methodology 255
  7.1 Antecedents and related work ................................ 256
  7.2 The FS-SLRNN linear model of FSM representation in RNNs ............ 259
    7.2.1 The representation of the state transition function ............. 260
    7.2.2 Stability of state representation ............................ 262
    7.2.3 The representation of the output function ..................... 265
  7.3 Implementation of FSMs in RNNs using the FS-SLRNN model ............. 267
    7.3.1 Implementation of FSMs in first-order ASLRNNs .................. 267
    7.3.2 Implementation of FSMs in second- and higher-order SLRNNs 273
    7.3.3 Implementation of DFAs and stochastic DFAs .................... 275
    7.3.4 The problem of searching minimal size networks that implement a given FSM ................................ 277
8 Augmented regular expressions: a formalism to describe and recognize a class of context-sensitive languages

8.1 Obtaining regular expressions from FSAs

8.1.1 The original Arden's algorithm

8.1.2 A modification of Arden's algorithm to discriminate loops from other circuits in the resulting RE

8.1.3 Simplifying the regular expression obtained by Arden's algorithm

8.2 Augmented regular expressions (AREs)

8.2.1 The star variables and the star tree of a regular expression

8.2.2 The star instances data structure

8.2.3 Definition of ARE and language described by an ARE

8.3 Expressive power of AREs

8.4 String recognition through AREs

8.4.1 Parsing strings by REs to build the star instances structure

8.4.1.1 An algorithm to determine the star instances based on parsing by an RE alone

8.4.1.2 An algorithm to determine the star instances based on parsing by an RE with the help of an equivalent DFA

8.4.2 Testing the satisfaction of ARE constraints by star instances

9 Inductive inference of augmented regular expressions
9.1 A general approach to infer AREs from string examples ............... 346
9.2 Induction of ARE constraints from recorded star instances .......... 355
9.3 Experimental assessment of a specific method for inferring AREs from positive examples ............................................. 361
   9.3.1 A method for inferring AREs from positive examples ............. 361
   9.3.2 Experimental assessment .............................................. 362

10 Conclusions and future research ........................................ 367
   10.1 Summary of main results and conclusions .......................... 368
      10.1.1 Symbolic methods for regular grammatical inference .......... 368
      10.1.2 Connectionist and hybrid methods for regular grammatical inference ................................................................. 370
      10.1.3 Augmented regular expressions and their inductive inference .. 376
   10.2 Applications and future research ................................... 378
   10.3 List of major contributions ............................................ 380

References and Bibliography ............................................. 383