

Alex Graves

---

Supervised Sequence Labelling with Recurrent Neural Networks

# Studies in Computational Intelligence, Volume 385

## Editor-in-Chief

Prof. Janusz Kacprzyk  
Systems Research Institute  
Polish Academy of Sciences  
ul. Newelska 6  
01-447 Warsaw  
Poland  
E-mail: kacprzyk@ibspan.waw.pl

---

Further volumes of this series can be found on our homepage: [springer.com](http://springer.com)

Vol. 362. Pascal Bouvry, Horacio González-Vélez, and Joanna Kolodziej (Eds.)  
*Intelligent Decision Systems in Large-Scale Distributed Environments*, 2011  
ISBN 978-3-642-21270-3

Vol. 363. Kishan G. Mehrotra, Chilukuri Mohan, Jae C. Oh, Pramod K. Varshney, and Moonis Ali (Eds.)  
*Developing Concepts in Applied Intelligence*, 2011  
ISBN 978-3-642-21331-1

Vol. 364. Roger Lee (Ed.)  
*Computer and Information Science*, 2011  
ISBN 978-3-642-21377-9

Vol. 365. Roger Lee (Ed.)  
*Computers, Networks, Systems, and Industrial Engineering 2011*, 2011  
ISBN 978-3-642-21374-8

Vol. 366. Mario Köppen, Gerald Schaefer, and Ajith Abraham (Eds.)  
*Intelligent Computational Optimization in Engineering*, 2011  
ISBN 978-3-642-21704-3

Vol. 367. Gabriel Luque and Enrique Alba  
*Parallel Genetic Algorithms*, 2011  
ISBN 978-3-642-22083-8

Vol. 368. Roger Lee (Ed.)  
*Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing 2011*, 2011  
ISBN 978-3-642-22287-0

Vol. 369. Dominik Ryzko, Piotr Gawrysiak, Henryk Rybinski, and Marzena Kryszkiewicz (Eds.)  
*Emerging Intelligent Technologies in Industry*, 2011  
ISBN 978-3-642-22731-8

Vol. 370. Alexander Mehler, Kai-Uwe Kühnberger, Henning Lobin, Harald Lungen, Angelika Storrer, and Andreas Witt (Eds.)  
*Modeling, Learning, and Processing of Text Technological Data Structures*, 2011  
ISBN 978-3-642-22612-0

Vol. 371. Leonid Perlovsky, Ross Deming, and Roman Ilin (Eds.)  
*Emotional Cognitive Neural Algorithms with Engineering Applications*, 2011  
ISBN 978-3-642-22829-2

Vol. 372. António E. Ruano and Annamária R. Várkonyi-Kóczy (Eds.)  
*New Advances in Intelligent Signal Processing*, 2011  
ISBN 978-3-642-11738-1

Vol. 373. Oleg Okun, Giorgio Valentini, and Matteo Re (Eds.)  
*Ensembles in Machine Learning Applications*, 2011  
ISBN 978-3-642-22909-1

Vol. 374. Dimitri Plemenos and Georgios Miaoulis (Eds.)  
*Intelligent Computer Graphics 2011*, 2011  
ISBN 978-3-642-22906-0

Vol. 375. Marenglen Biba and Fatos Xhafa (Eds.)  
*Learning Structure and Schemas from Documents*, 2011  
ISBN 978-3-642-22912-1

Vol. 376. Toyohide Watanabe and Lakhmi C. Jain (Eds.)  
*Innovations in Intelligent Machines – 2*, 2012  
ISBN 978-3-642-23189-6

Vol. 377. Roger Lee (Ed.)  
*Software Engineering Research, Management and Applications 2011*, 2011  
ISBN 978-3-642-23201-5

Vol. 378. János Fodor, Ryszard Klempous, and Carmen Paz Suárez Araujo (Eds.)  
*Recent Advances in Intelligent Engineering Systems*, 2011  
ISBN 978-3-642-23228-2

Vol. 379. Ferrante Neri, Carlos Cotta, and Pablo Moscato (Eds.)  
*Handbook of Memetic Algorithms*, 2012  
ISBN 978-3-642-23246-6

Vol. 380. Anthony Brabazon, Michael O'Neill, and Dietmar Maringer (Eds.)  
*Natural Computing in Computational Finance*, 2011  
ISBN 978-3-642-23335-7

Vol. 381. Paweł Katarzyniak, Tzu-Fu Chiu, Chao-Fu Hong, and Ngoc Thanh Nguyen (Eds.)  
*Semantic Methods for Knowledge Management and Communication*, 2011  
ISBN 978-3-642-23417-0

Vol. 382. F.M.T. Brazier, Kees Nieuwenhuis, Gregor Pavlin, Martijn Warnier, and Costin Badica (Eds.)  
*Intelligent Distributed Computing V*, 2011  
ISBN 978-3-642-24012-6

Vol. 383. Takayuki Ito, Minjie Zhang, Valentin Robu, Shaheen Fatima, and Tokuro Matsuo (Eds.)  
*New Trends in Agent-based Complex Automated Negotiations*, 2011  
ISBN 978-3-642-24695-1

Vol. 384. Daphna Weinsshall, Jörn Anemüller, and Luc van Gool (Eds.)  
*Detection and Identification of Rare Audiovisual Cues*, 2012  
ISBN 978-3-642-24033-1

Vol. 385. Alex Graves  
*Supervised Sequence Labelling with Recurrent Neural Networks*, 2012  
ISBN 978-3-642-24796-5

Alex Graves

# Supervised Sequence Labelling with Recurrent Neural Networks

*Author*

Dr. Alex Graves  
University of Toronto  
Department of Computer Science  
Toronto, Ontario  
Canada

ISSN 1860-949X

e-ISSN 1860-9503

ISBN 978-3-642-24796-5

e-ISBN 978-3-642-24797-2

DOI 10.1007/978-3-642-24797-2

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2011940011

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Contents

<b>List of Tables</b>	<b>IX</b>
<b>List of Figures</b>	<b>XI</b>
<b>List of Algorithms</b>	<b>XIII</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Structure of the Book . . . . .	3
<b>2 Supervised Sequence Labelling</b>	<b>5</b>
2.1 Supervised Learning . . . . .	5
2.2 Pattern Classification . . . . .	6
2.2.1 Probabilistic Classification . . . . .	6
2.2.2 Training Probabilistic Classifiers . . . . .	7
2.2.3 Generative and Discriminative Methods . . . . .	8
2.3 Sequence Labelling . . . . .	9
2.3.1 Sequence Classification . . . . .	11
2.3.2 Segment Classification . . . . .	11
2.3.3 Temporal Classification . . . . .	12
<b>3 Neural Networks</b>	<b>15</b>
3.1 Multilayer Perceptrons . . . . .	15
3.1.1 Forward Pass . . . . .	16
3.1.2 Output Layers . . . . .	18
3.1.3 Loss Functions . . . . .	19
3.1.4 Backward Pass . . . . .	20
3.2 Recurrent Neural Networks . . . . .	22
3.2.1 Forward Pass . . . . .	22
3.2.2 Backward Pass . . . . .	23
3.2.3 Unfolding . . . . .	24
3.2.4 Bidirectional Networks . . . . .	24
3.2.5 Sequential Jacobian . . . . .	27
3.3 Network Training . . . . .	29
3.3.1 Gradient Descent Algorithms . . . . .	29
3.3.2 Generalisation . . . . .	30

3.3.3	Input Representation . . . . .	34
3.3.4	Weight Initialisation . . . . .	35
<b>4</b>	<b>Long Short-Term Memory</b>	<b>37</b>
4.1	Network Architecture . . . . .	37
4.2	Influence of Preprocessing . . . . .	41
4.3	Gradient Calculation . . . . .	42
4.4	Architectural Variants . . . . .	42
4.5	Bidirectional Long Short-Term Memory . . . . .	43
4.6	Network Equations . . . . .	43
4.6.1	Forward Pass . . . . .	44
4.6.2	Backward Pass . . . . .	44
<b>5</b>	<b>A Comparison of Network Architectures</b>	<b>47</b>
5.1	Experimental Setup . . . . .	47
5.2	Network Architectures . . . . .	48
5.2.1	Computational Complexity . . . . .	49
5.2.2	Range of Context . . . . .	49
5.2.3	Output Layers . . . . .	49
5.3	Network Training . . . . .	51
5.3.1	Retraining . . . . .	51
5.4	Results . . . . .	52
5.4.1	Previous Work . . . . .	54
5.4.2	Effect of Increased Context . . . . .	54
5.4.3	Weighted Error . . . . .	56
<b>6</b>	<b>Hidden Markov Model Hybrids</b>	<b>57</b>
6.1	Background . . . . .	57
6.2	Experiment: Phoneme Recognition . . . . .	58
6.2.1	Experimental Setup . . . . .	59
6.2.2	Results . . . . .	59
<b>7</b>	<b>Connectionist Temporal Classification</b>	<b>61</b>
7.1	Background . . . . .	61
7.2	From Outputs to Labellings . . . . .	63
7.2.1	Role of the Blank Labels . . . . .	64
7.2.2	Bidirectional and Unidirectional Networks . . . . .	64
7.3	Forward-Backward Algorithm . . . . .	64
7.3.1	Log Scale . . . . .	68
7.4	Loss Function . . . . .	68
7.4.1	Loss Gradient . . . . .	69
7.5	Decoding . . . . .	71
7.5.1	Best Path Decoding . . . . .	71
7.5.2	Prefix Search Decoding . . . . .	72
7.5.3	Constrained Decoding . . . . .	73
7.6	Experiments . . . . .	78

7.6.1	Phoneme Recognition 1 . . . . .	79
7.6.2	Phoneme Recognition 2 . . . . .	80
7.6.3	Keyword Spotting . . . . .	81
7.6.4	Online Handwriting Recognition . . . . .	85
7.6.5	Offline Handwriting Recognition . . . . .	88
7.7	Discussion . . . . .	91
<b>8</b>	<b>Multidimensional Networks</b>	<b>95</b>
8.1	Background . . . . .	95
8.2	Network Architecture . . . . .	97
8.2.1	Multidirectional Networks . . . . .	100
8.2.2	Multidimensional Long Short-Term Memory . . . . .	102
8.3	Experiments . . . . .	104
8.3.1	Air Freight Data . . . . .	104
8.3.2	MNIST Data . . . . .	105
8.3.3	Analysis . . . . .	107
<b>9</b>	<b>Hierarchical Subsampling Networks</b>	<b>109</b>
9.1	Network Architecture . . . . .	110
9.1.1	Subsampling Window Sizes . . . . .	112
9.1.2	Hidden Layer Sizes . . . . .	112
9.1.3	Number of Levels . . . . .	113
9.1.4	Multidimensional Networks . . . . .	113
9.1.5	Output Layers . . . . .	115
9.1.6	Complete System . . . . .	116
9.2	Experiments . . . . .	117
9.2.1	Offline Arabic Handwriting Recognition . . . . .	119
9.2.2	Online Arabic Handwriting Recognition . . . . .	122
9.2.3	French Handwriting Recognition . . . . .	124
9.2.4	Farsi/Arabic Character Classification . . . . .	126
9.2.5	Phoneme Recognition . . . . .	127
	<b>References</b>	<b>133</b>
	<b>Acknowledgements</b>	<b>141</b>

# List of Tables

5.1	Framewise phoneme classification results on TIMIT . . . . .	52
5.2	Comparison of BLSTM with previous network . . . . .	54
6.1	Phoneme recognition results on TIMIT . . . . .	60
7.1	Phoneme recognition results on TIMIT with 61 phonemes . .	79
7.2	Folding the 61 phonemes in TIMIT onto 39 categories . . . .	80
7.3	Phoneme recognition results on TIMIT with 39 phonemes . .	82
7.4	Keyword spotting results on Verbmobil . . . . .	83
7.5	Character recognition results on IAM-OnDB . . . . .	86
7.6	Word recognition on IAM-OnDB . . . . .	88
7.7	Word recognition results on IAM-DB . . . . .	92
8.1	Classification results on MNIST . . . . .	106
9.1	Networks for offline Arabic handwriting recognition . . . . .	121
9.2	Offline Arabic handwriting recognition competition results . .	121
9.3	Networks for online Arabic handwriting recognition . . . . .	124
9.4	Online Arabic handwriting recognition competition results . .	125
9.5	Network for French handwriting recognition . . . . .	126
9.6	French handwriting recognition competition results . . . . .	127
9.7	Networks for Farsi/Arabic handwriting recognition . . . . .	129
9.8	Farsi/Arabic handwriting recognition competition results . .	129
9.9	Networks for phoneme recognition on TIMIT . . . . .	129
9.10	Phoneme recognition results on TIMIT . . . . .	130



# List of Figures

2.1	Sequence labelling . . . . .	9
2.2	Three classes of sequence labelling task . . . . .	10
2.3	Importance of context in segment classification . . . . .	12
3.1	A multilayer perceptron . . . . .	16
3.2	Neural network activation functions . . . . .	17
3.3	A recurrent neural network . . . . .	22
3.4	An unfolded recurrent network . . . . .	25
3.5	An unfolded bidirectional network . . . . .	26
3.6	Sequential Jacobian for a bidirectional network . . . . .	28
3.7	Overfitting on training data . . . . .	31
3.8	Different Kinds of Input Perturbation . . . . .	33
4.1	The vanishing gradient problem for RNNs . . . . .	38
4.2	LSTM memory block with one cell . . . . .	39
4.3	An LSTM network . . . . .	40
4.4	Preservation of gradient information by LSTM . . . . .	41
5.1	Various networks classifying an excerpt from TIMIT . . . . .	50
5.2	Framewise phoneme classification results on TIMIT . . . . .	53
5.3	Learning curves on TIMIT . . . . .	53
5.4	BLSTM network classifying the utterance “one oh five” . . . . .	55
7.1	CTC and framewise classification . . . . .	62
7.2	Unidirectional and Bidirectional CTC Networks Phonetically Transcribing an Excerpt from TIMIT . . . . .	65
7.3	CTC forward-backward algorithm . . . . .	67
7.4	Evolution of the CTC error signal during training . . . . .	70
7.5	Problem with best path decoding . . . . .	71
7.6	Prefix search decoding . . . . .	72
7.7	CTC outputs for keyword spotting on Verbmobil . . . . .	84
7.8	Sequential Jacobian for keyword spotting on Verbmobil . . . . .	84
7.9	BLSTM-CTC network labelling an excerpt from IAM-OnDB . . . . .	87

7.10	BLSTM-CTC Sequential Jacobian from IAM-OnDB with raw inputs . . . . .	89
7.11	BLSTM-CTC Sequential Jacobian from IAM-OnDB with preprocessed inputs . . . . .	90
8.1	MDRNN forward pass . . . . .	97
8.2	MDRNN backward pass . . . . .	97
8.3	Sequence ordering of 2D data . . . . .	98
8.4	Context available to a unidirectional two dimensional RNN . . . . .	100
8.5	Axes used by the hidden layers in a multidirectional MDRNN . . . . .	101
8.6	Context available to a multidirectional MDRNN . . . . .	101
8.7	Frame from the Air Freight database . . . . .	105
8.8	MNIST image before and after deformation . . . . .	106
8.9	MDRNN applied to an image from the Air Freight database . . . . .	107
8.10	Sequential Jacobian of an MDRNN for an image from MNIST . . . . .	107
9.1	Information flow through an HSRNN . . . . .	110
9.2	An unfolded HSRNN . . . . .	111
9.3	Information flow through a multidirectional HSRNN . . . . .	114
9.4	HSRNN applied to offline Arabic handwriting recognition . . . . .	118
9.5	Offline Arabic word images . . . . .	120
9.6	Offline Arabic error curves . . . . .	122
9.7	Online Arabic input sequences . . . . .	124
9.8	French word images . . . . .	125
9.9	Farsi character images . . . . .	128
9.10	Three representations of a TIMIT utterance . . . . .	130

# List of Algorithms

3.1	BRNN Forward Pass . . . . .	26
3.2	BRNN Backward Pass . . . . .	26
3.3	Online Learning with Gradient Descent . . . . .	30
3.4	Online Learning with Gradient Descent and Weight Noise . .	33
7.1	Prefix Search Decoding . . . . .	74
7.2	CTC Token Passing . . . . .	77
8.1	MDRNN Forward Pass . . . . .	99
8.2	MDRNN Backward Pass . . . . .	99
8.3	Multidirectional MDRNN Forward Pass . . . . .	102
8.4	Multidirectional MDRNN Backward Pass . . . . .	102