

References

- Abed, H.E., Margner, V., Kherallah, M., Alimi, A.M.: Online Arabic Handwriting Recognition Competition. In: 10th International Conference on Document Analysis and Recognition, ICDAR 2009, pp. 1388–1392. IEEE Computer Society (2009)
- An, G.: The Effects of Adding Noise During Backpropagation Training on a Generalization Performance. *Neural Computation* 8(3), 643–674 (1996) ISSN 0899-7667
- Bakker, B.: Reinforcement Learning with Long Short-Term Memory. In: *Advances in Neural Information Processing Systems*, vol. 14 (2002)
- Baldi, P., Pollastri, G.: The Principled Design of Large-scale Recursive Neural Network Architectures—DAG-RNNs and the Protein Structure Prediction Problem. *The Journal of Machine Learning Research* 4, 575–602 (2003) ISSN 1533-7928
- Baldi, P., Brunak, S., Frasconi, P., Soda, G., Pollastri, G.: Exploiting the Past and the Future in Protein Secondary Structure Prediction. *Bioinformatics* 15 (1999)
- Baldi, P., Brunak, S., Frasconi, P., Pollastri, G., Soda, G.: Bidirectional Dynamics for Protein Secondary Structure Prediction. In: Sun, R., Giles, C.L. (eds.) *Sequence Learning*. LNCS (LNAI), vol. 1828, pp. 80–104. Springer, Heidelberg (2001)
- Bayer, J., Wierstra, D., Togelius, J., Schmidhuber, J.: Evolving Memory Cell Structures for Sequence Learning. In: Alippi, C., Polycarpou, M., Panayiotou, C., Ellinas, G. (eds.) *ICANN 2009*. LNCS, vol. 5769, pp. 755–764. Springer, Heidelberg (2009)
- Bengio, Y.: A Connectionist Approach to Speech Recognition. *International Journal on Pattern Recognition and Artificial Intelligence* 7(4), 647–668 (1993)
- Bengio, Y.: Markovian Models for Sequential Data. *Neural Computing Surveys* 2, 129–162 (1999)
- Bengio, Y., LeCun, Y.: Scaling learning algorithms towards AI. In: Bottou, L., Chapelle, O., DeCoste, D., Weston, J. (eds.) *Large-Scale Kernel Machines*. MIT Press (2007)
- Bengio, Y., De Mori, R., Flammia, G., Kompe, R.: Global Optimization of a Neural Network–Hidden Markov Model Hybrid. *IEEE Transactions on Neural Networks* 3(2), 252–259 (1992)
- Bengio, Y., Simard, P., Frasconi, P.: Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Networks* 5(2), 157–166 (1994)
- Bengio, Y., LeCun, Y., Nohl, C., Burges, C.: LeRec: A NN/HMM Hybrid for On-line Handwriting Recognition. *Neural Computation* 7(6), 1289–1303 (1995)
- Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H.: Greedy Layer-wise Training of Deep Networks. In: Schölkopf, B., Platt, J., Hoffman, T. (eds.) *Advances in Neural Information Processing Systems*, vol. 19, pp. 153–160. MIT Press, Cambridge (2007)

- Beringer, N.: Human Language Acquisition in a Machine Learning Task. In: International Conference on Spoken Language Processing (2004)
- Bertolami, R., Bunke, H.: Multiple Classifier Methods for Offline Handwritten Text Line Recognition. In: 7th International Workshop on Multiple Classifier Systems, Prague, Czech Republic (2007)
- Bishop, C.M.: *Neural Networks for Pattern Recognition*. Oxford University Press (1995)
- Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer, Heidelberg (2006)
- Bottou, L., LeCun, Y.: Graph Transformer Networks for Image Recognition. In: Proceedings of ISI (2005)
- Bourlard, H., Morgan, N.: *Connectionist Speech Recognition: A Hybrid Approach*. Kluwer Academic Publishers (1994)
- Bourlard, H., Konig, Y., Morgan, N., Ris, C.: A new training algorithm for hybrid HMM/ANN speech recognition systems. In: 8th European Signal Processing Conference, vol. 1, pp. 101–104 (1996)
- Bridle, J.S.: Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition. In: Fogelman-Soulie, F., Herault, J. (eds.) *Neurocomputing: Algorithms, Architectures and Applications*, pp. 227–236. Springer, Heidelberg (1990)
- Broomhead, D., Lowe, D.: Multivariate Functional Interpolation and Adaptive Networks. *Complex Systems* 2, 321–355 (1988)
- Byrd, R.H., Lu, P., Nocedal, J., Zhu, C.Y.: A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing* 16(6), 1190–1208 (1995)
- Chang, J.: *Near-Miss Modeling: A Segment-Based Approach to Speech Recognition*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (1998)
- Chen, J., Chaudhari, N.: Protein Secondary Structure Prediction with bidirectional LSTM networks. In: International Joint Conference on Neural Networks: Post-Conference Workshop on Computational Intelligence Approaches for the Analysis of Bio-data (CI-BIO) (August 2005)
- Chen, J., Chaudhari, N.S.: Capturing Long-term Dependencies for Protein Secondary Structure Prediction. In: Yin, F.-L., Wang, J., Guo, C. (eds.) *ISNN 2004, Part II. LNCS*, vol. 3174, pp. 494–500. Springer, Heidelberg (2004)
- Chen, R., Jamieson, L.: Experiments on the Implementation of Recurrent Neural Networks for Speech Phone Recognition. In: Proceedings of the Thirtieth Annual Asilomar Conference on Signals, Systems and Computers, pp. 779–782 (1996)
- Decoste, D., Schölkopf, B.: Training Invariant Support Vector Machines. *Machine Learning* 46(1-3), 161–190 (2002)
- Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*. Wiley-Interscience Publication (2000)
- Eck, D., Schmidhuber, J.: Finding Temporal Structure in Music: Blues Improvisation with LSTM Recurrent Networks. In: Bourlard, H. (ed.) *Proceedings of the 2002 IEEE Workshop on Neural Networks for Signal Processing XII*, pp. 747–756. IEEE, New York (2002)
- Elman, J.L.: Finding Structure in Time. *Cognitive Science* 14, 179–211 (1990)
- Fahlman, S.: Faster Learning Variations on Back-propagation: An Empirical Study. In: Touretzky, D., Hinton, G., Sejnowski, T. (eds.) *Proceedings of the 1988 Connectionist Models Summer School*, pp. 38–51. Morgan Kaufmann (1989)

- Fernández, S., Graves, A., Schmidhuber, J.: An Application of Recurrent Neural Networks to Discriminative Keyword Spotting. In: de Sá, J.M., Alexandre, L.A., Duch, W., Mandic, D.P. (eds.) ICANN 2007. LNCS, vol. 4669, pp. 220–229. Springer, Heidelberg (2007)
- Fernandez, S., Graves, A., Schmidhuber, J.: Phoneme Recognition in TIMIT with BLSTM-CTC. Technical Report IDSIA-04-08, IDSIA (April 2008)
- Frasconi, P., Gori, M., Sperduti, A.: A General Framework for Adaptive Processing of Data Structures. *IEEE Transactions on Neural Networks* 9, 768–786 (1998)
- Fukada, T., Schuster, M., Sagisaka, Y.: Phoneme Boundary Estimation Using Bidirectional Recurrent Neural Networks and its Applications. *Systems and Computers in Japan* 30(4), 20–30 (1999)
- Garofolo, J.S., Lamel, L.F., Fisher, W.M., Fiscus, J.G., Pallett, D.S., Dahlgren, N.L.: DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus CDROM (1993)
- Gers, F.: Long Short-Term Memory in Recurrent Neural Networks. PhD thesis, Ecole Polytechnique Fédérale de Lausanne (2001)
- Gers, F., Schraudolph, N., Schmidhuber, J.: Learning Precise Timing with LSTM Recurrent Networks. *Journal of Machine Learning Research* 3, 115–143 (2002)
- Gers, F.A., Schmidhuber, J.: LSTM Recurrent Networks Learn Simple Context Free and Context Sensitive Languages. *IEEE Transactions on Neural Networks* 12(6), 1333–1340 (2001)
- Gers, F.A., Schmidhuber, J., Cummins, F.: Learning to Forget: Continual Prediction with LSTM. *Neural Computation* 12(10), 2451–2471 (2000)
- Giraud-Carrier, C., Vilalta, R., Brazdil, P.: Introduction to the Special Issue on Meta-Learning. *Machine Learning* 54(3), 187–193 (2004)
- Glass, J.R.: A Probabilistic Framework for Segment-based Speech Recognition. *Computer Speech and Language* 17, 137–152 (2003)
- Goller, C.: A Connectionist Approach for Learning Search-Control Heuristics for Automated Deduction Systems. PhD thesis, Fakultät für Informatik der Technischen Universität München (1997)
- Graves, A., Schmidhuber, J.: Framewise Phoneme Classification with Bidirectional LSTM Networks. In: Proceedings of the 2005 International Joint Conference on Neural Networks (2005a)
- Graves, A., Schmidhuber, J.: Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures. *Neural Networks* 18(5-6), 602–610 (2005b)
- Graves, A., Schmidhuber, J.: Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks. In: Koller, D., Schuurmans, D., Bengio, Y., Bottou, L. (eds.) *Advances in Neural Information Processing Systems*, vol. 21, pp. 545–552. MIT Press (2009)
- Graves, A., Beringer, N., Schmidhuber, J.: Rapid Retraining on Speech Data with LSTM Recurrent Networks. Technical Report IDSIA-09-05, IDSIA (2005a)
- Graves, A., Fernández, S., Schmidhuber, J.: Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition. In: Duch, W., Kacprzyk, J., Oja, E., Zdrożny, S. (eds.) ICANN 2005. LNCS, vol. 3697, pp. 799–804. Springer, Heidelberg (2005)
- Graves, A., Fernández, S., Gomez, F., Schmidhuber, J.: Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks. In: Proceedings of the International Conference on Machine Learning, ICML 2006, Pittsburgh, USA (2006)
- Graves, A., Fernández, S., Schmidhuber, J.: Multi-dimensional Recurrent Neural Networks. In: de Sá, J.M., Alexandre, L.A., Duch, W., Mandic, D.P. (eds.) ICANN 2007. LNCS, vol. 4668, pp. 549–558. Springer, Heidelberg (2007)

- Graves, A., Fernández, S., Liwicki, M., Bunke, H., Schmidhuber, J.: Unconstrained Online Handwriting Recognition with Recurrent Neural Networks. In: Platt, J., Koller, D., Singer, Y., Roweis, S. (eds.) *Advances in Neural Information Processing Systems*, vol. 20. MIT Press, Cambridge (2008)
- Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., Schmidhuber, J.: A Novel Connectionist System for Unconstrained Handwriting Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(5), 855–868 (2009)
- Grosicki, E., Abed, H.E.: ICDAR 2009 Handwriting Recognition Competition. In: 10th International Conference on Document Analysis and Recognition, pp. 1398–1402 (2009)
- Grosicki, E., Carre, M., Brodin, J.-M., Geoffrois, E.: Results of the RIMES Evaluation Campaign for Handwritten Mail Processing. In: International Conference on Document Analysis and Recognition, pp. 941–945 (2009)
- Halberstadt, A.K.: Heterogeneous Acoustic Measurements and Multiple Classifiers for Speech Recognition. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (1998)
- Hammer, B.: On the Approximation Capability of Recurrent Neural Networks. *Neurocomputing* 31(1–4), 107–123 (2000)
- Hammer, B.: Recurrent Networks for Structured Data - a Unifying Approach and Its Properties. *Cognitive Systems Research* 3, 145–165 (2002)
- Hennebert, J., Ris, C., Bourlard, H., Renals, S., Morgan, N.: Estimation of Global Posteriors and Forward-backward Training of Hybrid HMM/ANN Systems. In: Proc. of the European Conference on Speech Communication and Technology (Eurospeech 1997), pp. 1951–1954 (1997)
- Hestenes, M.R., Stiefel, E.: Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards* 49(6), 409–436 (1952)
- Hifny, Y., Renals, S.: Speech Recognition using Augmented Conditional Random Fields. *Trans. Audio, Speech and Lang. Proc.* 17, 354–365 (2009)
- Hinton, G.E., van Camp, D.: Keeping Neural Networks Simple by Minimizing the Description Length of the Weights. In: Conference on Learning Theory, pp. 5–13 (1993)
- Hinton, G.E., Osindero, S., Teh, Y.-W.: A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation* 18(7), 1527–1554 (2006)
- Hochreiter, S.: Untersuchungen zu Dynamischen Neuronalen Netzen. PhD thesis, Institut für Informatik, Technische Universität München (1991)
- Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. *Neural Computation* 9(8), 1735–1780 (1997)
- Hochreiter, S., Bengio, Y., Frasconi, P., Schmidhuber, J.: Gradient Flow in Recurrent Nets: the Difficulty of Learning Long-term Dependencies. In: Kremer, S.C., Kolen, J.F. (eds.) *A Field Guide to Dynamical Recurrent Neural Networks*. IEEE Press (2001a)
- Hochreiter, S., Bengio, Y., Frasconi, P., Schmidhuber, J.: Gradient flow in Recurrent Nets: the Difficulty of Learning Long-term Dependencies. In: Kremer, S.C., Kolen, J.F. (eds.) *A Field Guide to Dynamical Recurrent Neural Networks*. IEEE Press (2001b)
- Hochreiter, S., Heusel, M., Obermayer, K.: Fast Model-based Protein Homology Detection without Alignment. *Bioinformatics* (2007)
- Hopfield, J.J.: Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences of the United States of America* 79(8), 2554–2558 (1982)

- Hornik, K., Stinchcombe, M., White, H.: Multilayer Feedforward Networks are Universal Approximators. *Neural Networks* 2(5), 359–366 (1989)
- Hülksen, F., Wallhoff, F., Rigoll, G.: Facial Expression Recognition with Pseudo-3D Hidden Markov Models. In: Radig, B., Florczyk, S. (eds.) *DAGM 2001*. LNCS, vol. 2191, pp. 291–297. Springer, Heidelberg (2001)
- Jaeger, H.: The “Echo State” Approach to Analysing and Training Recurrent Neural Networks. Technical Report GMD Report 148, German National Research Center for Information Technology (2001)
- Jim, K.-C., Giles, C., Horne, B.: An Analysis of Noise in Recurrent Neural Networks: Convergence and Generalization. *IEEE Transactions on Neural Networks* 7(6), 1424–1438 (1996)
- Jiten, J., Mérialdo, B., Huet, B.: Multi-dimensional Dependency-tree Hidden Markov Models. In: *International Conference on Acoustics, Speech, and Signal Processing* (2006)
- Johansson, S., Atwell, R., Garside, R., Leech, G.: The tagged LOB corpus user’s manual; Norwegian Computing Centre for the Humanities (1986)
- Johnson, M.T.: Capacity and Complexity of HMM Duration Modeling techniques. *IEEE Signal Processing Letters* 12(5), 407–410 (2005)
- Jordan, M.I.: Attractor dynamics and parallelism in a connectionist sequential machine, pp. 112–127. *IEEE Press* (1990)
- Joshi, D., Li, J., Wang, J.: Parameter Estimation of Multi-dimensional Hidden Markov Models: A Scalable Approach. In: *Proc. of the IEEE International Conference on Image Processing (ICIP 2005)*, pp. 149–152 (2005)
- Kadous, M.W.: Temporal Classification: Extending the Classification Paradigm to Multivariate Time Series. PhD thesis, School of Computer Science & Engineering, University of New South Wales (2002)
- Kershaw, D., Robinson, A., Hochberg, M.: Context-Dependent Classes in a Hybrid Recurrent Network-HMM Speech Recognition System. In: Touretzky, D.S., Mozer, M.C., Hasselmo, M.E. (eds.) *Advances in Neural Information Processing Systems*, vol. 8, pp. 750–756. MIT Press (1996)
- Khosravi, H., Kabir, E.: Introducing a Very Large Dataset of Handwritten Farsi Digits and a Study on their Varieties. *Pattern Recogn. Lett.* 28, 1133–1141 (2007)
- Kohonen, T.: *Self-organization and Associative Memory*, 3rd edn. Springer, New York (1989)
- Koistinen, P., Holmström, L.: Kernel Regression and Backpropagation Training with Noise. In: Moody, J.E., Hanson, S.J., Lippmann, R. (eds.) *Advances in Neural Information Processing Systems*, vol. 4, pp. 1033–1039. Morgan Kaufmann (1991)
- Lafferty, J.D., McCallum, A., Pereira, F.C.N.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: *Proceedings of the Eighteenth International Conference on Machine Learning, ICML 2001*, pp. 282–289. Morgan Kaufmann Publishers Inc. (2001)
- Lamel, L., Gauvain, J.: High Performance Speaker-Independent Phone Recognition Using CDHMM. In: *Proc. Eurospeech* (September 1993)
- Lang, K.J., Waibel, A.H., Hinton, G.E.: A Time-delay Neural Network Architecture for Isolated Word Recognition. *Neural Networks* 3(1), 23–43 (1990)
- LeCun, Y., Bottou, L., Bengio, Y.: Reading Checks with Graph Transformer Networks. In: *International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 151–154. *IEEE* (1997)
- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE* 86(11), 2278–2324 (1998a)

- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 1–46 (1998b)
- LeCun, Y.A., Bottou, L., Orr, G.B., Müller, K.-R.: Efficient BackProp. In: Orr, G.B., Müller, K.-R. (eds.) *NIPS-WS 1996*. LNCS, vol. 1524, pp. 9–50. Springer, Heidelberg (1998)
- Lee, K.-F., Hon, H.-W.: Speaker-independent Phone Recognition Using Hidden Markov Models. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37(11), 1641–1648 (1989)
- Li, J., Najmi, A., Gray, R.M.: Image Classification by a Two-Dimensional Hidden Markov Model. *IEEE Transactions on Signal Processing* 48(2), 517–533 (2000)
- Lin, T., Horne, B.G., Tiño, P., Giles, C.L.: Learning Long-Term Dependencies in NARX Recurrent Neural Networks. *IEEE Transactions on Neural Networks* 7(6), 1329–1338 (1996)
- Lindblad, T., Kinser, J.M.: *Image Processing Using Pulse-Coupled Neural Networks*. Springer-Verlag New York, Inc. (2005)
- Liwicki, M., Bunke, H.: Handwriting Recognition of Whiteboard Notes. In: *Proc. 12th Conf. of the International Graphonomics Society*, pp. 118–122 (2005a)
- Liwicki, M., Bunke, H.: IAM-OnDB - an On-Line English Sentence Database Acquired from Handwritten Text on a Whiteboard. In: *Proc. 8th Int. Conf. on Document Analysis and Recognition*, vol. 2, pp. 956–961 (2005b)
- Liwicki, M., Graves, A., Fernández, S., Bunke, H., Schmidhuber, J.: A Novel Approach to On-Line Handwriting Recognition Based on Bidirectional Long Short-Term Memory Networks. In: *Proceedings of the 9th International Conference on Document Analysis and Recognition, ICDAR 2007* (September 2007)
- MacKay, D.J.C.: Probable Networks and Plausible Predictions - a Review of Practical Bayesian Methods for Supervised Neural Networks. *Network: Computation in Neural Systems* 6, 469–505 (1995)
- Märgner, V., Abed, H.E.: ICDAR 2009 Arabic Handwriting Recognition Competition. In: *10th International Conference on Document Analysis and Recognition*, pp. 1383–1387 (2009)
- Marti, U.-V., Bunke, H.: Using a Statistical Language Model to Improve the Performance of an HMM-based Cursive Handwriting Recognition System. *Int. Journal of Pattern Recognition and Artificial Intelligence* 15, 65–90 (2001)
- Marti, U.-V., Bunke, H.: The IAM Database: An English Sentence Database for Offline Handwriting Recognition. *International Journal on Document Analysis and Recognition* 5, 39–46 (2002)
- McCarter, G., Storkey, A.: *Air Freight Image Segmentation Database* (2007)
- McCulloch, W.S., Pitts, W.: A Logical Calculus of the Ideas Immanent in Nervous Activity, pp. 15–27. MIT Press (1988)
- Ming, J., Smith, F.J.: Improved Phone Recognition Using Bayesian Triphone Models. In: *ICASSP*, vol. 1, pp. 409–412 (1998)
- Mohamed, A., Dahl, G., Hinton, G.: Acoustic Modeling using Deep Belief Networks. *IEEE Transactions on Audio, Speech, and Language Processing* (99) (2011)
- Morris, J., Lussier, E.F.: Combining Phonetic Attributes Using Conditional Random Fields. In: *Proc. Interspeech 2006* (2006)
- Mozaffari, S., Faez, K., Faradji, F., Ziaratban, M., Golzan, S.M.: Comprehensive Isolated Farsi/Arabic Character Database for Handwritten OCR Research. In: Lorette, G. (ed.) *Tenth International Workshop on Frontiers in Handwriting Recognition*. Suvisoft (2006)
- Mozer, M.C.: Induction of Multiscale Temporal Structure. In: Moody, J.E., Hanson, S.J., Lippmann, R.P. (eds.) *Advances in Neural Information Processing Systems*, vol. 4, pp. 275–282. Morgan Kaufmann Publishers (1992)

- Murray, A.F., Edwards, P.J.: Enhanced MLP Performance and Fault Tolerance Resulting from Synaptic Weight Noise During Training. *IEEE Transactions on Neural Networks* 5, 792–802 (1994)
- Navarro, G.: A guided tour to approximate string matching. *ACM Computing Surveys* 33(1), 31–88 (2001)
- Neal, R.M.: *Bayesian Learning for Neural Networks*. Springer, New York (1996)
- Neto, J., Almeida, L., Hochberg, M., Martins, C., Nunes, L., Renals, S., Robinson, A.: Speaker Adaptation for Hybrid HMM-ANN Continuous Speech Recognition System. In: *Proceedings of Eurospeech 1995*, vol. 1, pp. 2171–2174 (1995)
- Pang, X., Werbos, P.J.: Neural Network Design for J Function Approximation in Dynamic Programming. *Mathematical Modeling and Scientific Computing* 5(2/3) (1996)
- Pechwitz, M., Maddouri, S.S., Mrgner, V., Ellouze, N., Amiri, H.: IFN/ENIT - Database of Handwritten Arabic Words. In: *Colloque International Francophone sur l'Écrit et le Document*, pp. 129–136 (2002)
- Plate, T.A.: Holographic Recurrent Networks. In: Giles, C.L., Hanson, S.J., Cowan, J.D. (eds.) *Advances in Neural Information Processing Systems*, vol. 5, pp. 34–41. Morgan Kaufmann (1993)
- Plaut, D.C., Nowlan, S.J., Hinton, G.E.: Experiments on Learning by Back-Propagation. Technical Report CMU-CS-86-126, Carnegie-Mellon University (1986)
- Pollastri, G., Vullo, A., Frasconi, P., Baldi, P.: Modular DAG-RNN Architectures for Assembling Coarse Protein Structures. *Journal of Computational Biology* 13(3), 631–650 (2006)
- Rabiner, L.R.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE* 77(2), 257–286 (1989)
- Reisenhuber, M., Poggio, T.: Hierarchical Models of Object Recognition in Cortex. *Nature Neuroscience* 2(11), 1019–1025 (1999)
- Renals, S., Morgan, N., Bourlard, H., Cohen, M., Franco, H.: Connectionist Probability Estimators in HMM Speech Recognition. *IEEE Transactions Speech and Audio Processing* (1993)
- Riedmiller, M., Braun, H.: A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP algorithm. In: *Proc. of the IEEE Intl. Conf. on Neural Networks*, San Francisco, CA, pp. 586–591 (1993)
- Robinson, A., Holdsworth, J., Patterson, J., Fallside, F.: A comparison of preprocessors for the cambridge recurrent error propagation network speech recognition system. In: *Proceedings of the First International Conference on Spoken Language Processing, ICSLP 1990* (1990)
- Robinson, A.J.: Several Improvements to a Recurrent Error Propagation Network Phone Recognition System. Technical Report CUED/F-INFENG/TR82, University of Cambridge (1991)
- Robinson, A.J.: An Application of Recurrent Nets to Phone Probability Estimation. *IEEE Transactions on Neural Networks* 5(2), 298–305 (1994)
- Robinson, A.J., Fallside, F.: The Utility Driven Dynamic Error Propagation Network. Technical Report CUED/F-INFENG/TR.1, Cambridge University Engineering Department (1987)
- Robinson, A.J., Almeida, L., Boite, J.-M., Bourlard, H., Fallside, F., Hochberg, M., Kershaw, D., Kohn, P., Konig, Y., Morgan, N., Neto, J.P., Renals, S., Saerens, M., Wooters, C.: A Neural Network Based, Speaker Independent, Large Vocabulary, Continuous Speech Recognition System: the Wernicke Project. In: *Proc. of the Third European Conference on Speech Communication and Technology (Eurospeech 1993)*, pp. 1941–1944 (1993)

- Rosenblatt, F.: The Perceptron: a Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review* 65, 386–408 (1958)
- Rosenblatt, F.: *Principles of Neurodynamics*. Spartan, New York (1963)
- Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning Internal Representations by Error Propagation, pp. 318–362. MIT Press (1986)
- Russell, S., Norvig, P.: *Artificial Intelligence: A Modern Approach*, 2nd edn. Prentice-Hall, Englewood Cliffs (2003)
- Sainath, T., Ramabhadran, B., Picheny, M.: An Exploration of Large Vocabulary Tools for Small Vocabulary Phonetic Recognition. In: *IEEE Workshop on Automatic Speech Recognition Understanding, ASRU 2009*, pp. 359–364 (2009)
- Schmidhuber, J.: Learning Complex Extended Sequences using the principle of history compression. *Neural Computing*, 234–242 (1992)
- Schmidhuber, J., Wierstra, D., Gagliolo, M., Gomez, F.: Training Recurrent Networks by Evolino. *Neural Computation* 19(3), 757–779 (2007)
- Schraudolph, N.: Fast Curvature Matrix-Vector Products for Second-Order Gradient Descent. *Neural Computation* 14(7), 1723–1738 (2002)
- Schuster, M.: On Supervised Learning from Sequential Data With Applications for Speech Recognition. PhD thesis, Nara Institute of Science and Technology, Kyoto, Japan (1999)
- Schuster, M., Paliwal, K.K.: Bidirectional Recurrent Neural Networks. *IEEE Transactions on Signal Processing* 45, 2673–2681 (1997)
- Senior, A., Robinson, A.J.: Forward-Backward Retraining of Recurrent Neural Networks. In: Touretzky, D.S., Mozer, M.C., Hasselmo, M.E. (eds.) *Advances in Neural Information Processing Systems*, vol. 8, pp. 743–749. The MIT Press (1996)
- Sha, F., Saul, L.K.: Large Margin Hidden Markov Models for Automatic Speech Recognition. In: *Advances in Neural Information Processing Systems*, pp. 1249–1256 (2006)
- Shewchuk, J.R.: An Introduction to the Conjugate Gradient Method Without the Agonizing Pain. Technical report, Carnegie Mellon University, Pittsburgh, PA, USA (1994)
- Simard, P.Y., Steinkraus, D., Platt, J.C.: Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. In: *ICDAR 2003: Proceedings of the Seventh International Conference on Document Analysis and Recognition*. IEEE Computer Society (2003)
- Solimanpour, F., Sadri, J., Suen, C.Y.: Standard Databases for Recognition of Handwritten Digits, Numerical Strings, Legal Amounts, Letters and Dates in Farsi Language. In: Lorette, G. (ed.) *Tenth International Workshop on Frontiers in Handwriting Recognition*, Suvisoft (October 2006)
- Sperduti, A., Starita, A.: Supervised Neural Networks for the Classification of Structures. *IEEE Transactions on Neural Networks* 8(3), 714–735 (1997)
- Thireou, T., Reczko, M.: Bidirectional Long Short-Term Memory Networks for Predicting the Subcellular Localization of Eukaryotic Proteins. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 4(3), 441–446 (2007)
- Tosi, S.: *Matplotlib for Python Developers*. Packt Publishing (2009)
- Trentin, E., Gori, M.: Robust combination of neural networks and hidden Markov models for speech recognition. *IEEE Transactions on Neural Networks* 14(6), 1519–1531 (2003)
- Vapnik, V.N.: *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc. (1995)
- Verbmobil. Database Version 2.3 (2004)

- Welch, P.: The Use of Fast Fourier Transform for the Estimation of Power Spectra: a Method Based on Time Averaging over Short, Modified Periodograms. *IEEE Transactions on Audio and Electroacoustics* 15(2), 70–73 (1967)
- Werbos, P.: Backpropagation Through Time: What It Does and How to Do It. *Proceedings of the IEEE* 78(10), 1550–1560 (1990)
- Werbos, P.J.: Generalization of Backpropagation with Application to a Recurrent Gas Market Model. *Neural Networks* 1 (1988)
- Wierstra, D., Gomez, F.J., Schmidhuber, J.: Modeling systems with internal state using evoluno. In: *GECCO 2005: Proceedings of the 2005 Conference on Genetic and Evolutionary Computation*, pp. 1795–1802. ACM Press (2005)
- Williams, R.J., Zipser, D.: Gradient-Based Learning Algorithms for Recurrent Networks and Their Computational Complexity. In: Chauvin, Y., Rumelhart, D.E. (eds.) *Back-Propagation: Theory, Architectures and Applications*, pp. 433–486. Lawrence Erlbaum Publishers (1995)
- Wu, L., Baldi, P.: A Scalable Machine Learning Approach to Go. In: Schalkopf, B., Platt, J., Hoffman, T. (eds.) *Advances in Neural Information Processing Systems*, vol. 19, pp. 1521–1528. MIT Press (2006)
- Young, S., Russell, N., Thornton, J.: Token Passing: A Simple Conceptual Model for Connected Speech Recognition Systems. Technical Report CUED/F-INFENG/TR38, Cambridge University Engineering Dept., Cambridge, UK (1989)
- Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., Woodland, P.: *The HTK Book*. Cambridge University Engineering Department, HTK version 3.4 edition (December 2006)
- Yu, D., Deng, L., Acero, A.: A Lattice Search Technique for a Long-Contextual-Span Hidden Trajectory Model of Speech. *Speech Communication* 48(9), 1214–1226 (2006)
- Zavaliagos, G., Austin, S., Makhoul, J., Schwartz, R.M.: A Hybrid Continuous Speech Recognition System Using Segmental Neural Nets with Hidden Markov Models. *International Journal of Pattern Recognition and Artificial Intelligence* 7(4), 949–963 (1993)
- Zimmermann, H.G., Grothmann, R., Schaefer, A.M., Tietz, C.: Identification and Forecasting of Large Dynamical Systems by Dynamical Consistent Neural Networks. In: Haykin, S., Principe, J., Sejnowski, T., McWhirter, J. (eds.) *New Directions in Statistical Signal Processing: From Systems to Brain*, pp. 203–242. MIT Press (2006a)
- Zimmermann, M., Chappelier, J.-C., Bunke, H.: Offline Grammar-based Recognition of Handwritten Sentences. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(5), 818–821 (2006b)

Acknowledgements

I would first like to thank my supervisor Jürgen Schmidhuber for his guidance and support throughout the Ph.D. thesis on which this book was based. I would also like to thank my co-authors Santiago Fernández, Nicole Beringer, Faustino Gomez and Douglas Eck, and everyone else I collaborated with at IDSIA and the Technical University of Munich, for making the stimulating and creative places to work. Thanks to Tom Schaul for proofreading an early draft of the book, and to Marcus Hutter for his assistance during Chapter 7. I am grateful to Marcus Liwicki, Horst Bunke and Roman Bertolami for their expert collaboration on handwriting recognition. A special mention to all my friends in Lugano, Munich and elsewhere who made the whole thing worth doing: Frederick Ducatelle, Matteo Gagliolo, Nikos Mutsanas, Ola Svensson, Daniil Ryabko, John Paul Walsh, Adrian Taruttis, Andreas Brandmaier, Christian Osendorfer, Thomas Rückstieß, Justin Bayer, Murray Dick, Luke Williams, John Lord, Sam Mungall, David Larsen and all the rest. But most of all I would like to thank my family, my wife Alison and my children Liam and Nina for being there when I needed them most.

Alex Graves is a Junior Fellow of the Canadian Institute for Advanced Research.