Approximate recursive Bayesian estimation of recurrent neural networks: on-line learning of synaptic weights, neuron activities and network structure

Branimir Todorović[1]

[1]Faculty of Sciences and Mathematics, University of Niš, branimirtodorovic@pmf.ni.ac.rs

We have derived algorithms for on-line training of the recurrent neural networks using approximate recursive Bayesian estimation of unknown probabilty density function of its state, represented as concatenated vector of synaptic weights and neuron activities. Joint estimation of synaptic weights and neuron activities generalizes the heuristic known as teacher forcing, which enables filtering out the noise from date during the training. Optimal solution of the recursive Bayesian estimation for recurrent neural natworks is intractable, due to the nonlinearty of the network dynamics, therfore approximate solutons have to be considered. We have derived a class of derivative free algorithms for on-line training of recurrent networks, using Stirlings interpolation formula and the Unscented transformation. For the case when the non-Gaussian (multi modal or heavy tailed) noise is present on training data, we have derived learning algorithms using Gaussian mixture as the approximation of probability density function of the RNN state. Finally, we have used statistics, recursively updated during sequential Bayesian estimation, to derive criteria for growing and pruning of synaptic connections and hidden neurons in recurrent neural networks. The performance of the proposed learning algorithms is demonstrated on problems of chaotic and nonstationary time series prediction, blind signal separation and deconvolution and dynamic system identification.

References

[1] S. J. Julier and J. K. Uhlmann, A new extension of the Kalman filter to nonlinear systems, in: Preoceedings of AeroSense - The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls, Orlando, FL, 1997.

[2] M. Nørgaard, N. K. Poulsen and O. Ravn, Advances in derivative free state estimation for nonlinear systems, Technical Report, IMM-REP-1998-15, Department of Mathematical Modelling, DTU, 2000.

[3] B. Todorović, M. Stanković and C. Moraga, Extended Kalman filter based adaptation of time-varying recurrent radial basis function networks structure, in: P. Sinčák, J. Vaščák and K. Hirota (eds.) Machine Intelligence: Quo Vadis? Advances in Fuzzy Systems-Applications and Theory **21**, World Scientific, 2014, 115–124.

[4] B. Todorović, M. Stanković and C. Moraga, Nonlinear Bayesian estimation of recurrent neural networks, in: Proceedings of ISDA 2004 - IEEE 4th International Conference on Intelligent Systems Design and Applications, Budapest, Hungary, August 26-28, 2014, 855–860.

[5] B. Todorović, M. Stanković and C. Moraga, Gaussian sum filters for recurrent neural networks training, in: Proceeding of NEUREL 2006 - Eight Seminar on Neural Network Applications in Electrical Engineering, Belgrade, Serbia, September 25-27, 2006, 53–58.

[6] B. Todorović, M. Stanković and C. Moraga, Derivative free training of recurrent neural networks - a comparison of algorithms and architectures, in: Proceedings of NCTA 2014 - The International Conference on Neural Computation Theory and Applications, part of IJCCI 2014, Rome, Italy, October 22-24, 2014, 76–84.

[7] B. Todorović, M. Stanković and C. Moraga, Recurrent neural networks training using derivative free nonlinear Bayesian filters, in: Computational Intelligence, Studies in Computational Intelligence **620**, 2015, 383–410.