ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning. Bruges (Belgium), 22-24 April 2009, d-side publi., ISBN 2-930307-09-9.

Recent advances in efficient learning of recurrent networks

Barbara Hammer[1], Benjamin Schrauwen[2], Jochen J. Steil[3],

1- Institute of Informatics - Clausthal University of Technology - Germany

2- Reservoir Lab - Ghent University - Belgium

3- Research Institute of Cognition and Robotics - University of Bielefeld - Germany

Abstract. Recurrent neural networks (RNNs) carry the promise of implementing efficient and biologically plausible signal processing. They both are optimally suited for a wide area of applications when dealing with spatiotemporal data or causalities and provide explanation of cognitive phenomena of the human brain. Recently, a few new fundamental paradigms connected to RNNs have been developed which allow insights into their potential for information processing. They also pave the way towards new efficient training algorithms which overcome the well-known problem of long-term dependencies. This tutorial gives an overview of this recent developments in efficient, biologically plausible recurrent information processing.

1 Introduction

Recurrent neural networks (RNNs), in particular Hopfield networks, were among the models which led to a renaissance of neural network research in the 90th after the rapid decrease of interest in neural network research caused by the exact mathematical proof of the limitations of the simple perceptron and extensions thereof by Minsky and Papert in the 70th. Convincing properties of the Hopfield model model include biological plausibility as associative memory, exact mathematical treatment of the capacities and dynamics in terms of an underlying cost function and associated stable states, and the possibility to apply the model to classical problems, such as the traveling salesperson problem.

Since then, a variety of very different recurrent neural networks with different dynamics and application areas has emerged. The models include partially recurrent systems such as discrete and continuous time models for time series prediction, simulation of dynamical systems, and processing of language. Fully recurrent models such as Boltzmann machines and associative memories, models involving only local or restricted recurrence such as long short term memory networks or locally recurrent, globally feedforward networks, recurrent models which resemble classical filters, or recurrent systems aimed for an explanation of biological recurrent networks such as the neocortex have also been developed.

When investigating these models, the key issues remain widely the same: the biological plausibility of recurrent systems and their connected dynamic phenomena, the exact mathematical characterization of the network dynamics, and applicability to relevant problems. While the mathematical treatment of recurrent networks of various form has reached a matured state in some respects, the applicability to classical problems and training of recurrent networks still faces severe problems. Practitioners therefore often prefer a simple feedforward modeling over the in theory much more powerful recurrent treatment of problems involving temporal or spatial dependencies.

We will argue in this overview that new developments in the context of recurrent models, in particular a matured characterization of stability of recurrent systems, new biologically plausible training modes, and network models such as reservoir computing have lead to a state of the art where these problems may be overcome.

2 Applications

One of the most general formulations of the dynamics of recurrent networks is given by the simple formula

$$y(t+1) = f(x(t), y(t), o(t)), \quad o(t+1) = g(y(t))$$

in discrete time or

$$\dot{y} \sim f(x, y, o), \quad \dot{o} \sim g(y)$$

in continuous time with inputs x, hidden states y, outputs o, and nonlinear activation functions f and g. These formulas can be used to model different dynamics: on the one hand the most common usage of RNNs is short or long term prediction of time series, when only a fixed time horizon is predicted based on given inputs and an initial state. On the other hand, they model also approximation of global fixed points when the system is driven up to convergence based on a given initial state or, more generally, generation of patterns when the system is driven towards an autonomously generated periodic, quasi-periodic, or more complex (possibly chaotic) attractor.

Correspondingly, recent applications of recurrent networks can be found in a variety of different areas. One of the classical application scenarios of RNNs is the prediction of time series and classification of temporal patterns [16, 47] such as the prediction of the stock market [68], of electricity load [10], the modeling of natural phenomena such as sandbars or wind speed [100, 6, 5], the classification of birdsongs [59], speech or speaker recognition [60, 34], event detection in robotics [2], or the modeling of speech and grammar [18, 32]. When a RNN is put into a loop with the system nonlinear control can be achieved. This is one of the most popular application areas of RNNs, accompanied by a vast literature concerning control designs which fulfill certain important properties such as stability or optimality [83, 76, 104, 46, 155, 130, 128]. Applications in this context reach from the control of industrial and technical systems [110, 104, 62] up to robotics [154, 53]. When referring to stable states of RNNs or their long term dynamics in the form of associative memories, RNNs can be used to solve complex dynamic tasks such as classical problems including the traveling salesperson problem [134, 125], Boolean factor analysis [33], or object segmentation in images [90, 144]. Further, RNNs can be used to directly model biological counterparts such as biological networks; successful inference of gene regulatory networks by means of RNNs has been presented e.g. in [70, 149].

3 Cost function optimization

The original Hopfield network uses the RNN dynamics to retrieve stable states associated to initial conditions of a given network the their respective basins of attraction. One of the achievements of Hopfield was the explicit formulation of a corresponding cost function which is optimized by the network dynamics. This principle lies behind a variety of general design rules of RNNs to solve complex dynamical tasks. One of the early applications which are still investigated is given by solutions for the traveling salesperson problem [134, 125]. More general approaches design RNNs to solve general linear or quadratic optimization problems [50, 4, 23, 79, 80, 82] possibly with constraints [36, 146] which can be used e.g. for the distributed training of support vector machines [105]. Counterparts in continuous space concern various variational problems (linear/nonlinear, possibly with constraints) [147, 49, 145, 48]. Due to its biological plausibility, another very popular function designed by means of fully recurrent RNNs is the winner takes all function and generalizations thereof such as the k-WTA function, as investigated e.g. in [81, 15, 115]. Apart from applications e.g. for the rapid parallel retrieval of data from huge databases, these models are relevant to explain parts of the functionality of biological systems.

4 Stability

The attractors of an associative memory are the stable states of the underlying dynamics. The notion of stability, however, concerns a much richer mathematical terrain, as pioneered e.g. in the approaches [131, 128] for RNNs. While stable states or stable attractors constitute the essential solutions of associative memories and RNNs used for optimization purpose, stability constitutes a key property for RNNs used for control tasks. Thereby, different mathematical notions of stability exist, often global exponential stability (as strongest and, therefore, most desirable property) is proved, but weaker conditions are also investigated, including new notions of stability such as e.g. investigated in [22].

Much research has been done to find easy to test conditions for stability of RNNs with adaptive parameters (weights and delays) and different activation functions (assuming additional properties such as boundedness or differentiability, if necessary). Often, standard techniques from stability theory are applied such as Lyapunov functions, and, mostly, conditions in the form of linear matrix inequalities are derived, which can be tested in practice, see e.g. the articles [160, 121, 111, 126, 161, 156, 86, 123, 21, 78, 77, 1, 103, 158, 84, 157, 17, 73, 75, 130, 128] for stability properties under different conditions. Some approaches deal with the stabilization of RNNs [103, 122] e.g. by inducing appropriate noise into the system. Moreover, apart from stability conditions, analysis of the number of stable states or periodic attractors and the corresponding basins of attraction is of interest [86, 148, 74]. This knowledge is very relevant to design reliable learning algorithms for fully recurrent networks acting as associative memory, on the one hand, and RNNs used for control, on the other hand. Further, the dynamic properties of RNNs as manifested in their stable states are of relevance to fully understand RNNs which model biological systems.

5 Mathematical analysis

Closely related to stability analysis is the analysis of the dynamics of RNNs. This extends classical stability analysis e.g. considering conditions for multiperiodicity [162, 159] and an exact investigation of the dynamical properties of RNNs under different conditions. It is well known that the full spectrum of dynamic behavior can be observed already in very small networks. While small systems (e.g. three neurons) can still be analyzed analytically [35, 9, 44], the situation becomes more and more complex for larger networks. Correspondingly, approximations [72] or a more global analysis become necessary. For large (random) models, the effect of the network topology on important properties can provide such measurement [106], e.g. the ability of random networks to solve benchmark pattern recognition tasks [31], the effect of the topology on the synchronizability of such systems [19], or the capacity of such systems depending on topology and form of neurons [69, 138]. The approach [40] successfully derives a mapping of meta-parameters of RNNs derived similar to Google's PageRank to the resulting dynamics.

These findings are interesting for the design of training algorithms or the design of network blocks which can act e.g. as pattern generator. Further, when investigating biologically plausible models such as e.g. linear threshold circuits [133] or spiking networks [3], the findings can help to correctly interpret the function of biological systems. Very interesting results have also been found in the frame of language learning, where the state space reveals insights into the underlying recursive process [20, 135].

6 Biologically inspired models

The mathematical analysis of RNNs serves different purposes: on the one hand, it helps to design efficient training algorithms, on the other hand, it allows insight into the principles underlying RNNs. Thereby, the design and analysis of models which mirror important properties of biological systems is of particular relevance to the understanding of biological RNNs and, eventually, the mechanisms underlying the human brain. Research which investigates biological aspects of RNNs can be classified into different categories: Some approaches model biologically plausible RNNs at a very abstract level and try to understand fundamental properties of such network architectures. As an example, the approach [36] models large random RNNs and investigates synchronization in such systems, since synchronization can frequently be observed in biological systems and it might play an important role in information coding. Similarly, the article [66] investigates correlations which occur in large networks. The approach [96] also deals with the encoding of temporal stimuli in biologically plausible RNNs. Biologically relevant networks architectures including RNNs with lateral inhibition, circuits of leaky integrate and fire neurons, and systems of spiking neurons are investigated in [89, 114, 92, 67], and relevant properties such as mean firing rate and dynamical characteristics are systematically described. Further, some work deals with simplified descriptions and models such as e.g. [102], which presents a general possibility to substitute RNN synapses with mixed sign by biologically more plausible models involving only one sign, or [150] which describes connections of biological RNNs to classical Kalman filters. Interesting further properties concern the role of noise in biological systems, including the tolerance of stable oscillatory patterns to noise [13], or even potential beneficial effects of response noise to better account for general synaptic noise [8].

Other approaches deal directly with a specific biological system or biological effect and devise simple models which can explain such findings and, eventually, lead to a better understanding of potential underlying mechanisms. This includes, for example, RNN models of the hippocampus [71], of cerebellar learning [48], the visual cortex or retina [139, 98, 65], mirror neurons [24], as well as RNN

ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning. Bruges (Belgium), 22-24 April 2009, d-side publi., ISBN 2-930307-09-9.

models which mimic specific effects such as retrospective and prospective recall activity [63], visual aftereffects and illusory contour formation [28, 91], pattern generation of the locomotor system [51], the emergence of dendritic stimulus selectivity [93] or modeling of eye saccades [97]. These models are partially highly specific and this field is closely connected to biological research and new developments and techniques in neurosciences. It targets at complementing the theoretical models by experimental data.

Several particularly interesting research directions deal with the understanding and modeling of biologically plausible and effective training mechanisms implemented in biological networks. One very active and particularly promising field of research concerns biological explanations of reinforcement learning processes in the brain, as presented e.g. in [107, 45, 116]. Reinforcement learning can bridge the gap between fully unsupervised paradigms, which can be well explained by different biologically plausible unsupervised models developed already at a very early stage in neural network research e.g. by Kohonen and von der Malsburg (recent work on unsupervised organization in RNN can be found in [38]), and supervised learning, which often relies on biologically not very plausible mathematical terms such as gradients and higher order derivatives.

7 New training paradigms

One of the merits of a deeper mathematical analysis of RNNs is the beneficial effect on efficient and robust training algorithms. As an example, conditions for RNN stability can be directly put as constraints to RNN learning applied to control tasks. Traditionally, RNN training mostly takes place according to one of two different paradigms: extensions of Hebbian learning for fully recurrent auto-associative memory RNNs, and gradient based approaches for classical supervised tasks in the frame of time series modeling and prediction. For the computation of gradients, either backpropagation through time (BPTT) or real time recurrent learning (RTRL) is commonly used, constituting efficient ways to compute the gradients (and, if necessary, also higher derivatives) in a forward-backward or purely forward way. However, it has been observed very early that gradient based training faces the problem of long-term dependencies, i.e. numerically stable backpropagation of error signals over long periods of time is hardly possible. Correspondingly, a large variety of modifications has been proposed, but still no widely accepted efficient training model exists in this framework.

Research on RNN training separates into diverse approaches to tackle this fundamental problem: some work deals with a better numerical realization of basic RTRL and BPTT training, including e.g. adaptive learning rates and a combination of BPTT and RTRL [127, 85, 153, 25] as well as extensions of these classical paradigms to more general models such as complex valued or spiking neurons [136]. As an alternative training paradigm, extended Kalman filter training is frequently used for RNN training, showing good results depending on the underlying task at hand [26, 109, 37, 108]. Further, formulations of RNN training as a reinforcement problem become more and more common to overcome the problem of temporal credit assignment. However, these methods often suffer from very long training times [39, 132, 52]. Alternative optimization methods are offered by classical meta-heuristics which are adapted to RNN parameter optimization such as evolutionary algorithms or particle swarm optimization [83, 149, 117]. Some approaches try to avoid the problem of vanishing gradients by designing the architecture appropriately, e.g. restricting recurrence basically to linear neurons as done in long short term memory and extensions thereof [117].

If prior knowledge about the task is available, knowledge based modeling can shape the search space towards efficient solutions. One common framework is offered by neuro-fuzzy systems such as e.g. classical Mamdani or Takagi-Sugeno controllers put into the architecture of an appropriate RNN [95, 61]. Efficiency can also be gained by means of a decomposition of the learning task using e.g. mixture of experts [95, 64] or incremental approaches [99]. As an alternative to neuro-fuzzy systems, the close link of RNNs to the dynamics of classical finite automata has been observed already very early and methods to put known automata rules into a RNN before training have been proposed. The inverse process, extraction of finite automata from RNNs, is also approximately possible (since RNNs are more powerful than finite automata, this is necessary an approximation), and can lead to much better long term behavior and generalization than the original RNN [54].

Despite these promising approaches no efficient learning algorithm for RNNs which is universally accepted exists by now, and also the underlying theory which guarantees the principled learnability of RNNs e.g. in the classical sense of probably approximate learnability is far from understood [41, 94].

8 Reservoir computing

One of the fundamental new paradigms and possible breakthroughs in RNN learning is offered by the principle of reservoir computing, which has been introduced simultaneously in several RNN models and corresponding learning algorithms, in particular in echo state networks (ESN) [55] and liquid state machines (LSM) [88]. Later it was shown under the notion backpropagation decorrelation (BD) that a reservoir with online learning naturally arises by simplification of a rigid error function minimization approach [129]. Basically, this type RNN learning rules rely on a usually fixed recurrent network often termed “dynamic reservoir” or just “reservoir”, which displays a rich but untrained dynamics, and a simply trainable readout. This mostly consists of only one linear layer and can therefore be trained analytically using e.g. the standard pseudo-inverse of the reservoir activations or using standard online schemes like least mean squares. LSM, ESN, and BD originally differed in the exact form of the reservoirs and learning rules, but these techniques are now generally referred to as Reservoir Computing (RC) systems [143]. Recent theoretical work [118] even further brought the theoretical work on ESNs and LSMs closer together. Interestingly, these models are not only very fast to train, but they are also biologically relevant, since training can be restricted to direct Hebb training of the readout instead of a mathematically complicated gradient training of the recurrent reservoir. The overall structure and learning method is very similar to models of the cerebellum, a connection which has been explored to some detail in [152] .

Surprisingly, these models show excellent learning behavior depending on the task at hand (see e.g. [57]). It has been shown that they constitute universal approximators of relevant functions such as filters with fading memory. However, their full potential regarding practical applications as well as ways to best design the reservoir for a given task have not yet been fully understood.

Recent research in reservoir computing concerns, on the one hand, applications and benchmark comparisons to test the applicability of these ideas to different areas. This includes applications for language modeling [137], control [141], and time series modeling [142], for example. Optimal training of reservoir networks constitutes a further active field of research, whereby this concerns both, the adaptation of the readout as well as an optimum design of the recurrent reservoir (an overview can be found in [87]). While the readout relies on supervised paradigms including, for example, support vector machines [124] or logistic regression [29], the reservoir is created based on general principles or using only very slow adaptation. Promising approaches propose to modify the connections such that the states are Gaussian distributed [120] or Laplacian distributed [12], to use highly clustered connectivity as present e.g. in scale-free networks [27], to enlarge the reservoir capacity by means of leaky integrators [58], or to include stability issues into reservoir adaptation such that maximum decorrelation is achieved together with robustness [130]. Note that, while adapting the reservoir, its stability should always be guaranteed, which can be guaranteed by classical conditions of stability theory as presented e.g. in [130, 14].

While RC is usually applied to spatio-temporal or sequential input, in recent work, the use of RC on static or semi-static data has also been researched [30, 113]. In this case the reservoir acts as a primal-space kernel which is intrinsically regularized by the stable dynamics of the recurrent network. In this context, it is of particular interest that a reservoir based architecture can allow online learning of static mappings, while data are presented in temporal correlation [112]. This is a very typical task in robotics applications, where data are generated and evaluated along trajectories such that previously only local models with their limited generalization capabilities could be used.

An important recent trend is the research on hierarchical reservoir systems. One of the disadvantages of RC systems is that they have trouble coping with spatiotemporal data where the relevant information spans a large frequency range (such as in speech: ranging from sound features over phonemes to words). A first solution is presented in [56], where a hierarchical RC system is presented where the layered readout is trained using backpropagation. A fundamental work in this direction recently showed that such a hierarchically organized recurrent architecture can emerge also under a classical BPTT training scheme when using different time scales [151].

Interestingly, reservoir models constitute efficient biologically plausible models which also allow a very efficient realization in hardware as demonstrated e.g. in the approach [119] or even as a novel computing approach in photonics [140]. Although their practical applicability and optimal training is still a matter of recent research, they constitute one of the most promising RNN models of recent years to bridge the gap between biologically plausible models and efficient learning systems with non-trivial applications.

9 New models

Apart from reservoir models, quite a few other new RNN models have been proposed in recent years to fill the gap of spatiotemporal pattern processing using connectionist systems. Mostly, these new models can be put into the frame of classical RNNs as characterized by the standard dynamic equations above. However, a few models have been proposed which add fundamental new principles to recurrent processing rather than only a minor extension of known

ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning. Bruges (Belgium), 22-24 April 2009, d-side publi., ISBN 2-930307-09-9.

models. As already mentioned, based on the principle of full recurrence and cost function optimization, RNNs can solve difficult optimization tasks. One particularly interesting model is offered by the competitive layer model which can be used as an efficient and trainable RNN for perceptual grouping, thus automatically solving non-trivial tasks in image processing [144] and automatic segmentation of task demonstrations in imitation learning [101].

Partially recurrent networks are driven by the input sequence and can thus be used for time series prediction and modeling. This principle has been generalized to more complex recursive structures, in particular tree and graph structures: so-called recursive neural networks constitute well-established models for the processing of trees, thus they can be used to bridge the gap between symbolic data and connectionist processing, since trees naturally encode formulas and symbolic terms. These models have recently been generalized to graph structures including acyclic and even cyclic graphs, thus opening the way towards efficient and fault tolerant machine learning techniques on general graphs [43, 11].

One further active and emerging field of research is the unsupervised training of recursive systems, where very diverse architectures which mirror different principles and capacities exist [42, 7].

10 Conclusions

One of the next big challenges in machine learning will be the processing of temporal or sequential data. In many application areas such speech recognition, robot control, forecasting, language understanding, etc. the current approaches seem to asymptote to a level of performance well below what humans can achieve. RNNs form a very compelling candidate for solving this conundrum: the model has strong connections to biological systems, it has extensively been theoretically studied, and owing to the recent advances in learning methods, can be efficiently and robustly trained. The next hurdle is devising architectures and techniques that can cope with vast and noisy application domains where information present in a wide temporal range needs to be integrated. Online and incremental learning in real time will get even more important for adaptive technology like interacting robots or speech processing devices. A further issue concerns data with different intrinsic timescales as e.g. found in robotics, where controller and planning levels contribute differently to movement, which can be tackled by hierarchies of systems or more complex recurrent networks displaying different timescales as well. As discussed, to all of these problems first pioneering work is just starting and can now rely on a large variety of promising approaches.

References

[1] C. W. Anderson, P. M. Young, M. R. Buehner, J. N. Knight, K. A. Bush, and D. C. Hittle. Robust reinforcement learning control using integral quadratic constraints for recurrent neural networks. *IEEE Transactions on Neural Networks*, 18(4):993–1002, 2007.

[2] E. A. Antonelo, B. Schrauwen, and D. Stroobandt. Event detection and localization for small mobile robots using Reservoir Computing. *Neural Networks*, 21(6):862–871, 2008.

[3] A. Banerjee, P. Serie‘s, and A. Pouget. Dynamical constraints on using precise spike timing to compute in recurrent cortical networks. *Neural Computation*, 20(4):974–993, 2008.

[4] M. P. Barbarosou and N. G. Maratos. A nonfeasible gradient projection recurrent neural network for equality-constrained optimization problems. *IEEE Trans. on Neural Networks*, 19(10):1665–1677, 2008.

[5] T. G. Barbounis and J. B. Theocharis. Locally recurrent neural networks for long-term wind speed and power prediction. *Neurocomputing*, 69(4-6):466–496, 2006.

[6] T. G. Barbounis and J. B. Theocharis. A locally recurrent fuzzy neural network with application to the wind speed prediction using spatial correlation. *Neurocomputing*, 70(7-9):1525–1542, 2007.

[7] G. D. A. Barreto, A. F. R. Araújo, and S. C. Kremer. A taxonomy for spatiotemporal connectionist networks revisited: The unsupervised case. *Neural Computation*, 15(6):1255–1320, 2003.

[8] G. Basalyga and E. Salinas. When response variability increases neural network robustness to synaptic noise. *Neural Computation*, 18(6):1349–1379, 2006.

[9] R. D. Beer. Parameter space structure of continuous-time recurrent neural networks. *Neural Computation*, 18(12):3009–3051, 2006.

[10] D. Benaouda, F. Murtagh, J. . Starck, and O. Renaud. Wavelet-based nonlinear multiscale decomposition model for electricity load forecasting. *Neurocomputing*, 70(1-3):139–154, 2006.

[11] M. Bianchini, M. Gori, L. Sarti, and F. Scarselli. Recursive processing of cyclic graphs. *IEEE Transactions on Neural Networks*, 17(1):10–18, 2006.

[12] J. Boedecker, O. Obst, N. M. Mayer, and M. Asada. Studies on reservoir initialization and dynamics shaping in echo state networks. In *Proceedings of the European Symposium on Neural Networks (ESANN)*, 2009.

[13] N. Brunel and D. Hansel. How noise affects the synchronization properties of recurrent networks of inhibitory neurons. *Neural Computation*, 18(5):1066–1110, 2006.

[14] M. Buehner and P. Young. A tighter bound for the echo state property. *IEEE Transactions on Neural Networks*, 17(3):820–824, 2006.

[15] X. Cai, D. V. Prokhorov, and D. C. Wunsch II. Training winner-take-all simultaneous recurrent neural networks. *IEEE Transactions on Neural Networks*, 18(3):674–684, 2007.

[16] X. Cai, N. Zhang, G. K. Venayagamoorthy, and D. C. Wunsch II. Time series prediction with recurrent neural networks trained by a hybrid PSO-EA algorithm. *Neurocomputing*, 70(13-15):2342–2353, 2007.

[17] J. Cao, K. Yuan, and H. Li. Global asymptotical stability of recurrent neural networks with multiple discrete delays and distributed delays. *IEEE Transactions on Neural Networks*, 17(6):1646–1651, 2006.

[18] B. Cartling. On the implicit acquisition of a context-free grammar by a simple recurrent neural network. *Neurocomputing*, 71(7-9):1527–1537, 2008.

[19] C. Cebulla. Asymptotic behavior and synchronizability characteristics of a class of recurrent neural networks. *Neural Computation*, 19(9):2492–2514, 2007.

[20] M. Cernansky, M. Makula, and L. Benuskova. Organization of the state space of a simple recurrent network before and after training on recursive linguistic structures. *Neural Networks*, 20(2):236–244, 2007.

[21] B. Chen and J. Wang. Global exponential periodicity and global exponential stability of a class of recurrent neural networks with various activation functions and time-varying delays. *Neural Networks*, 20(10):1067–1080, 2007.

[22] T. Chen and L. Wang. Global μ-stability of delayed neural networks with unbounded time-varying delays. *IEEE Transactions on Neural Networks*, 18(6):1836–1840, 2007.

[23] G. Costantini, R. Perfetti, and M. Todisco. Quasi-Lagrangian neural network for convex quadratic optimization. *IEEE Transactions on Neural Networks*, 19(10):1804–1809, 2008.

[24] R. Cuijpers, F. Stuijt, and I. Sprinkhuizen-Kuyper. Generalisation of action sequences in rnnpb networks with mirror properties. In *Proceedings of the European Symposium on Neural Networks (ESANN)*, 2009.

[25] O. De Jesus and M. T. Hagan. Backpropagation algorithms for a broad class of dynamic networks. *IEEE Transactions on Neural Networks*, 18(1):14–27, 2007.

[26] J. de Jesus Rubio and W. Yu. Nonlinear system identification with recurrent neural networks and dead-zone Kalman filter algorithm. *Neurocomputing*, 70(13-15):2460–2466, 2007.

[27] Z. Deng and Y. Zhang. Collective behavior of a small-world recurrent neural system with scale-free distribution. *IEEE Transactions on Neural Networks*, 18(5):1364–1375, 2007.

[28] D. Domijan, M. Setic, and D. Svegar. A model of the illusory contour formation based on dendritic computation. *Neurocomputing*, 70(10-12):1977–1982, 2007.

[29] X. Dutoit, B. Schrauwen, and H. Van Brussel. Non-markovian process modelling with echo state networks. In *Proceedings of the European Symposium on Neural Networks (ESANN)*, 2009.

[30] M. Embrechts, L. Alexandre, and J. Linton. Reservoir computing for static pattern

recognition. In *Proceedings of the European Symposium on Neural Networks (ESANN)*, 2009.

[31] F. Emmert-Streib. Influence of the neural network topology on the learning dynamics. *Neurocomputing*, 69(10-12):1179–1182, 2006.

[32] I. Farkas and M. W. Crocker. Syntactic systematicity in sentence processing with a recurrent self-organizing network. *Neurocomputing*, 71(7-9):1172–1179, 2008.

[33] A. A. Frolov, D. Husek, I. P. Muraviev, and P. Y. Polyakov. Boolean factor analysis by attractor neural network. *IEEE Transactions on Neural Networks*, 18(3):698–707, 2007.

[34] T. D. Ganchev, D. K. Tasoulis, M. N. Vrahatis, and N. D. Fakotakis. Generalized locally recurrent probabilistic neural networks with application to text-independent speaker verification. *Neurocomputing*, 70(7-9):1424–1438, 2007.

[35] B. Gao and W. Zhang. Equilibria and their bifurcations in a recurrent neural network involving iterates of a transcendental function. *IEEE Transactions on Neural Networks*, 19(5):782–794, 2008.

[36] E. Gelenbe and S. Timotheou. Random neural networks with synchronized interactions. *Neural Computation*, 20(9):2308–2324, 2008.

[37] S. L. Goh and D. P. Mandic. An augmented extended Kalman filter algorithm for complex-valued recurrent neural networks. *Neural Computation*, 19(4):1039–1055, 2007.

[38] C. Gros and G. Kaczor. Stimulus processing and unsupervised learning in autonomously active recurrent networks. In *Proceedings of the European Symposium on Neural Networks (ESANN)*, 2009.

[39] A. Gruning. Elman backpropagation as reinforcement for simple recurrent networks. *Neural Computation*, 19(11):3108–3131, 2007.

[40] T. Gurel, L. De Raedt, and S. Rotter. Ranking neurons for mining structure-activity relations in biological neural networks: NeuronRank. *Neurocomputing*, 70(10-12):1897–1901, 2007.

[41] B. Hammer. Generalization ability of folding networks. *IEEE Transactions on Knowledge and Data Engineering*, 13(2):196–206, 2001.

[42] B. Hammer, A. Micheli, N. Neubauer, A. Sperduti, and M. Strickert. Self organizing maps for time series. In *Proceedings of WSOM 2005*, pages 115–122, 2005.

[43] B. Hammer, A. Micheli, and A. Sperduti. Adaptive contextual processing of structured data by recursive neural networks: A survey of computational properties. In B. Hammer and P. Hitzler, editors, *Perspectives of Neural-Symbolic Integration*, volume 77 of *Studies in computational Intelligence*, pages 67–94. Springer, Berlin, 2007.

[44] R. Haschke and J. J. Steil. Input space bifurcation manifolds of recurrent neural networks. *Neurocomputing*, 64C:25–38, 2005.

[45] F. Henry, E. Dauce, and H. Soula. Temporal pattern identification using spike-timing dependent plasticity. *Neurocomputing*, 70(10-12):2009–2016, 2007.

[46] Z. Hou, M. M. Gupta, P. N. Nikiforuk, M. Tan, and L. Cheng. A recurrent neural network for hierarchical control of interconnected dynamic systems. *IEEE Transactions on Neural Networks*, 18(2):466–481, 2007.

[47] X. Hu, D. V. Prokhorov, and D. C. Wunsch II. Time series prediction with a weighted bidirectional multi-stream extended Kalman filter. *Neurocomputing*, 70(13-15):2392–2399, 2007.

[48] X. Hu and J. Wang. Solving pseudomonotone variational inequalities and pseudoconvex optimization problems using the projection neural network. *IEEE Transactions on Neural Networks*, 17(6):1487–1499, 2006.

[49] X. Hu and J. Wang. Solving generally constrained generalized linear variational inequalities using the general projection neural networks. *IEEE Transactions on Neural Networks*, 18(6):1697–1708, 2007.

[50] X. Hu and J. Wang. An improved dual neural network for solving a class of quadratic programming problems and its k-winners-take-all application. *IEEE Tran. on Neural Networks*, 19(12):2022–2031, 2008.

[51] M. Huss and M. Rehn. Tonically driven and self-sustaining activity in the lamprey hemicord: When can they co-exist? *Neurocomputing*, 70(10-12):1882–1886, 2007.

[52] S. Ishii, M. Shidara, and K. Shibata. A model to explain the emergence of reward expectancy neurons using reinforcement learning and neural network. *Neurocomputing*, 69(10-12):1327–1331, 2006.

[53] M. Ito, K. Noda, Y. Hoshino, and J. Tani. Dynamic and interactive generation of object handling behaviors by a small humanoid robot using a dynamic neural network model. *Neural Networks*, 19(3):323–337, 2006.

[54] H. Jacobsson. The crystallizing substochastic sequential machine extractor: CrySSMEx. *Neural Computation*, 18(9):2211–2255, 2006.

[55] H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. Technical Report GMD Report 148, German National Research Center for In-

formation Technology, 2001.
[56] H. Jaeger. Discovering multiscale dynamical features with hierarchical Echo State Networks. Technical report, Jacobs University Bremen, 2007.
[57] H. Jaeger and H. Haas. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless telecommunication. *Science*, 308:78–80, 2004.
[58] H. Jaeger, M. Lukosevicius, D. Popovici, and U. Siewert. Optimization and applications of Echo State Networks with leaky-integrator neurons. *Neural Networks*, 20(3):335–352, 2007.
[59] C. Juang and T. Chen. Birdsong recognition using prediction-based recurrent neural fuzzy networks. *Neurocomputing*, 71(1-3):121–130, 2007.
[60] C. Juang, C. Chiou, and C. Lai. Hierarchical singleton-type recurrent neural fuzzy networks for noisy speech recognition. *IEEE Transactions on Neural Networks*, 18(3):833–843, 2007.
[61] C. Juang and I.-F. Chung. Recurrent fuzzy network design using hybrid evolutionary learning algorithms. *Neurocomputing*, 70(16-18):3001–3010, 2007.
[62] C. Juang, S. Huang, and F. Duh. Mold temperature control of a rubber injection-molding machine by TSK-type recurrent neural fuzzy network. *Neurocomputing*, 70(1-3):559–567, 2006.
[63] R. A. Koene and M. E. Hasselmo. Reversed and forward buffering of behavioral spike sequences enables retrospective and prospective retrieval in hippocampal regions CA3 and CA1. *Neural Networks*, 21(2-3):276–288, 2008.
[64] E. Kolman and M. Margaliot. A new approach to knowledge-based design of recurrent neural networks. *IEEE Transactions on Neural Networks*, 19(8):1389–1401, 2008.
[65] J. Kremkow, A. Kumar, S. Rotter, and A. Aertsen. Emergence of population synchrony in a layered network of the cat visual cortex. *Neurocomputing*, 70(10-12):2069–2073, 2007.
[66] B. Kriener, T. Tetzlaff, A. Aertsen, M. Diesmann, and S. Rotter. Correlations and population dynamics in cortical networks. *Neural Computation*, 20(9):2185–2226, 2008.
[67] A. Kumar, S. Schrader, A. Aertsen, and S. Rotter. The high-conductance state of cortical networks. *Neural Computation*, 20(1):1–43, 2008.
[68] Y. Kwon and B. Moon. A hybrid neurogenetic approach for stock forecasting. *IEEE Transactions on Neural Networks*, 18(3):851–864, 2007.
[69] A. Lazar, G. Pipa, and J. Triesch. Fading memory and time series prediction in recurrent networks with different forms of plasticity. *Neural Networks*, 20(3):312–322, 2007.
[70] W. Lee and K. Yang. A clustering-based approach for inferring recurrent neural networks as gene regulatory networks. *Neurocomputing*, 71(4-6):600–610, 2008.
[71] C. Leibold and R. Kempter. Memory capacity for sequences in a recurrent network with biological constraints. *Neural Computation*, 18(4):904–941, 2006.
[72] A. Levina, U. Ernst, and J. Michael Herrmann. Criticality of avalanche dynamics in adaptive recurrent networks. *Neurocomputing*, 70(10-12):1877–1881, 2007.
[73] P. Li and J. Cao. Stability in static delayed neural networks: A nonlinear measure approach. *Neurocomputing*, 69(13-15):1776–1781, 2006.
[74] T. Li and S. Fei. Exponential state estimation for recurrent neural networks with distributed delays. *Neurocomputing*, 71(1-3):428–438, 2007.
[75] X. Liao, Q. Luo, and Z. Zeng. Positive invariant and global exponential attractive sets of neural networks with time-varying delays. *Neurocomputing*, 71(4-6):513–518, 2008.
[76] C. Lin, L. Chen, and C. Chen. RCMAC hybrid control for MIMO uncertain nonlinear systems using sliding-mode technology. *IEEE Transactions on Neural Networks*, 18(3):708–720, 2007.
[77] M. Liu. Delayed standard neural network models for control systems. *IEEE Transactions on Neural Networks*, 18(5):1376–1391, 2007.
[78] P. Liu and Q. Han. Discrete-time analogs for a class of continuous-time recurrent neural networks. *IEEE Transactions on Neural Networks*, 18(5):1343–1355, 2007.
[79] Q. Liu and J. Wang. A one-layer recurrent neural network with a discontinuous activation function for linear programming. *Neural Computation*, 20(5):1366–1383, 2008.
[80] Q. Liu and J. Wang. A one-layer recurrent neural network with a discontinuous hard-limiting activation function for quadratic programming. *IEEE Transactions on Neural Networks*, 19(4):558–570, 2008.
[81] Q. Liu and J. Wang. Two k-winners-take-all networks with discontinuous activation functions. *Neural Networks*, 21(2-3):406–413, 2008.
[82] S. Liu and J. Wang. A simplified dual neural network for quadratic programming with its KWTA application. *IEEE Transactions on Neural Networks*, 17(6):1500–1510, 2006.
[83] S. Liu, Y. Wang, and Q. Zhu. Development of a new EDRNN procedure in control of

human arm trajectories. *Neurocomputing*, 72(1-3):490–499, 2008.

[84] Y. Liu, Z. Wang, and X. Liu. Global exponential stability of generalized recurrent neural networks with discrete and distributed delays. *Neural Networks*, 19(5):667–675, 2006.

[85] Z. Liu and I. Elhanany. A fast and scalable recurrent neural network based on stochastic meta descent. *IEEE Trans. on Neural Networks*, 19(9):1651–1657, 2008.

[86] X. Lou and B. Cui. Delay-dependent criteria for global robust periodicity of uncertain switched recurrent neural networks with time-varying delay. *IEEE Transactions on Neural Networks*, 19(4):549–557, 2008.

[87] M. Lukosevicius and H. Jaeger. Overview of reservoir recipes. Technical Report no. 11, Jacobs University Bremen, 2007.

[88] W. Maass, T. Natschläger, and H. Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–2560, 2002.

[89] Z. Mao and S. G. Massaquoi. Dynamics of winner-take-all competition in recurrent neural networks with lateral inhibition. *IEEE Transactions on Neural Networks*, 18(1):55–69, 2007.

[90] J. Martinez-Cantos, E. Carmona, A. Fernandez-Caballero, and M. T. Lopez. Parametric improvement of lateral interaction in accumulative computation in motion-based segmentation. *Neurocomputing*, 71(4-6):776–786, 2008.

[91] F. Menghini, N. van Rijsbergen, and A. Treves. Modelling adaptation aftereffects in associative memory. *Neurocomputing*, 70(10-12):2000–2004, 2007.

[92] S. Mikula and E. Niebur. Exact solutions for rate and synchrony in recurrent networks of coincidence detectors. *Neural Computation*, 20(11):2637–2661, 2008.

[93] K. Morita, M. Okada, and K. Aihara. Selectivity and stability via dendritic nonlinearity. *Neural Computation*, 19(7):1798–1853, 2007.

[94] M.Vidyasagar. *The theory of learning and generalization.* Springer, 2 edition, 2002.

[95] J. Namikawa and J. Tani. A model for learning to segment temporal sequences, utilizing a mixture of RNN experts together with adaptive variance. *Neural Networks*, 21(10):1466–1475, 2008.

[96] R. Natarajan, Q. J. M. Huys, P. Dayan, and R. S. Zemel. Encoding and decoding spikes for dynamic stimuli. *Neural Computation*, 20(9):2325–2360, 2008.

[97] M. Nikitchenko and A. Koulakov. Neural integrator: A sandpile model. *Neural Computation*, 20(10):2379–2417, 2008.

[98] W. Niu and J. Yuan. Recurrent network simulations of two types of non-concentric retinal ganglion cells. *Neurocomputing*, 70(13-15):2576–2580, 2007.

[99] H. Ohta and Y. P. Gunji. Recurrent neural network architecture with pre-synaptic inhibition for incremental learning. *Neural Networks*, 19(8):1106–1119, 2006.

[100] L. Pape, B. G. Ruessink, M. A. Wiering, and I. L. Turner. Recurrent neural network modeling of nearshore sandbar behavior. *Neural Networks*, 20(4):509–518, 2007.

[101] M. Pardowitz, R. Haschke, J. J. Steil, and H. Ritter. Gestalt-based action segmentation for robot task learning. In *Proc. IEEE Humanoids*, 2008.

[102] C. Parisien, C. H. Anderson, and C. Eliasmith. Solving the problem of negative synaptic weights in cortical models. *Neural Computation*, 20(6):1473–1494, 2008.

[103] K. Patan. Stability analysis and the stabilization of a class of discrete-time dynamic neural networks. *IEEE Transactions on Neural Networks*, 18(3):660–673, 2007.

[104] Y. Peng and C. Lin. Adaptive recurrent cerebellar model articulation controller for linear ultrasonic motor with optimal learning rates. *Neurocomputing*, 70(16-18):2626–2637, 2007.

[105] R. Perfetti and E. Ricci. Analog neural network for support vector machine learning. *IEEE Transactions on Neural Networks*, 17(4):1085–1091, 2006.

[106] F. Piekniewski and T. Schreiber. Spontaneous scale-free structure of spike flow graphs in recurrent neural networks. *Neural Networks*, 21(10):1530–1536, 2008.

[107] A. Ponzi. Dynamical model of salience gated working memory, action selection and reinforcement based on basal ganglia and dopamine feedback. *Neural Networks*, 21(2-3):322–330, 2008.

[108] D. V. Prokhorov. Training recurrent neurocontrollers for robustness with derivative-free Kalman filter. *IEEE Transactions on Neural Networks*, 17(6):1606–1616, 2006.

[109] D. V. Prokhorov. Training recurrent neurocontrollers for real-time applications. *IEEE Transactions on Neural Networks*, 18(4):1003–1015, 2007.

[110] D. V. Prokhorov. Toyota Prius HEV neurocontrol and diagnostics. *Neural Networks*, 21(2-3):458–465, 2008.

[111] H. Qu, Z. Yi, and X. Wang. Switching analysis of 2-D neural networks with nonsaturating linear threshold transfer functions. *Neurocomputing*, 72(1-3):413–419, 2008.

[112] F. R. Reinhart and J. J. Steil. Recurrent neural associative learning of forward and inverse kinematics for movement generation of the redundant pa-10 robot. In *Proc. LAB-RS*, volume 1, pages 35–40, 2008.

[113] R. F. Reinhart and J. J. Steil. Attractor-based computation with reservoirs for online learning of inverse kinematics. In *Proceedings of the European Symposium on Neural Networks (ESANN)*, 2009.

[114] A. Renart, R. Moreno-Bote, X. . Wang, and N. Parga. Mean-driven and fluctuation-driven persistent activity in recurrent networks. *Neural Computation*, 19(1):1–46, 2007.

[115] W. Richards, H. Sebastian Seung, and G. Pickard. Neural voting machines. *Neural Networks*, 19(8):1161–1167, 2006.

[116] Y. Sakai, H. Okamoto, and T. Fukai. Computational algorithms and neuronal network models underlying decision processes. *Neural Networks*, 19(8):1091–1105, 2006.

[117] J. Schmidhuber, D. Wierstra, M. Gagliolo, and F. Gomez. Training recurrent networks by evolino. *Neural Computation*, 19(3):757–779, 2007.

[118] B. Schrauwen, L. Buesing, and R. Legenstein. On computational power and the order-chaos phase transition in reservoir computing. In *Proceedings of NIPS 2008*, 2009.

[119] B. Schrauwen, M. D'Haene, D. Verstraeten, and J. V. Campenhout. Compact hardware Liquid State Machines on FPGA for real-time speech recognition. *Neural Networks*, 21(2-3):511–523, 2008.

[120] B. Schrauwen, M. Wardermann, D. Verstraeten, J. J. Steil, and D. Stroobandt. Improving reservoirs using intrinsic plasticity. *Neurocomputing*, 71(7-9):1159–1171, 2008.

[121] H. Shao. Delay-dependent stability for recurrent neural networks with time-varying delays. *IEEE Transactions on Neural Networks*, 19(9):1647–1651, 2008.

[122] Y. Shen and J. Wang. Noise-induced stabilization of the recurrent neural networks with mixed time-varying delays and Markovian-switching parameters. *IEEE Transactions on Neural Networks*, 18(6):1857–1862, 2007.

[123] Y. Shen and J. Wang. An improved algebraic criterion for global exponential stability of recurrent neural networks with time-varying delays. *IEEE Transactions on Neural Networks*, 19(3):528–531, 2008.

[124] Z. Shi and M. Han. Support vector echo-state machine for chaotic time-series prediction. *IEEE Transactions on Neural Networks*, 18(2):359–372, 2007.

[125] P. H. Siqueira, M. T. A. Steiner, and S. Scheer. A new approach to solve the traveling salesman problem. *Neurocomputing*, 70(4-6):1013–1021, 2007.

[126] Q. Song. Exponential stability of recurrent neural networks with both time-varying delays and general activation functions via LMI approach. *Neurocomputing*, 71(13-15):2823–2830, 2008.

[127] Q. Song, Y. Wu, and Y. C. Soh. Robust adaptive gradient-descent training algorithm for recurrent neural networks in discrete time domain. *IEEE Trans. on Neural Networks*, 19(11):1841–1853, 2008.

[128] J. J. Steil. *Input-Output Stability of Recurrent Neural Networks*. Cuvillier Verlag, Göttingen, 1999.

[129] J. J. Steil. Backpropagation-Decorrelation: Online recurrent learning with O(N) complexity. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, volume 1, pages 843–848, 2004.

[130] J. J. Steil. Online stability of backpropagation-decorrelation recurrent learning. *Neurocomputing*, 69(7-9 SPEC. ISS.):642–650, 2006.

[131] J. A. K. Suykens, B. D. Moor, and J. Vandewalle. Robust local stability of multilayer recurrent neural networks. *IEEE Trans. Neural Networks*, 11(1):222–229, 2000.

[132] I. Szita and A. Lorincz. PIRANHA: Policy iteration for recurrent artificial neural networks with hidden activities. *Neurocomputing*, 70(1-3):577–591, 2006.

[133] H. Tang, K. C. Tan, and E. J. Teoh. Dynamics analysis and analog associative memory of networks with LT neurons. *IEEE Transactions on Neural Networks*, 17(2):409–418, 2006.

[134] E. J. Teoh, K. C. Tan, H. J. Tang, C. Xiang, and C. K. Goh. An asynchronous recurrent linear threshold network approach to solving the traveling salesman problem. *Neurocomputing*, 71(7-9):1359–1372, 2008.

[135] P. Tino, B. Hammer, and M. Boden. Markovian bias of neural-based architectures. In B. Hammer and P. Hitzler, editors, *Perspectives of Neural-Symbolic Integration*, volume 77 of *Studies in computational Intelligence*, pages 95–134. Springer, Berlin, 2007.

[136] P. Tino and A. J. S. Mills. Learning beyond finite memory in recurrent networks of spiking neurons. *Neural Computation*, 18(3):591–613, 2006.

[137] M. H. Tong, A. D. Bickett, E. M. Christiansen, and G. W. Cottrell. Learning grammatical structure with Echo State Networks. *Neural Networks*, 20(3):424–432, 2007.

[138] I. Y. Tyukin, D. Prokhorov, and C. Van Leeuwen. Adaptive classification of temporal

signals in fixed-weight recurrent neural networks: An existence proof. *Neural Computation*, 20(10):2564–2596, 2008.

[139] M. C. W. Van Rossum, M. A. A. Van Der Meer, D. Xiao, and M. W. Oram. Adaptive integration in the visual cortex by depressing recurrent cortical circuits. *Neural Computation*, 20(7):1847–1872, 2008.

[140] K. Vandoorne, W. Dierckx, B. Schrauwen, D. Verstraeten, R. Baets, P. Bienstman, and J. Van Campenhout. Toward optical signal processing using photonic reservoir computing. *Optics Express*, 16(15):11182–11192, 8 2008.

[141] G. K. Venayagamoorthy. Online design of an Echo State Network based wide area monitor for a multimachine power system. *Neural Networks*, 20(3):404–413, 2007.

[142] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt. An experimental unification of Reservoir Computing methods. *Neural Networks*, 20(3):391–403, 2007.

[143] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt. A unifying comparison of reservoir computing methods. *Neural Networks*, 20:391–403, 2007.

[144] S. Weng, H. Wersing, J. J. Steil, and H. Ritter. Learning lateral interactions for feature binding and sensory segmentation from prototypic basis interactions. *IEEE Transactions on Neural Networks*, 17(4):843–862, 2006.

[145] Y. Xia and G. Feng. A new neural network for solving nonlinear projection equations. *Neural Networks*, 20(5):577–589, 2007.

[146] Y. Xia, G. Feng, and J. Wang. A novel recurrent neural network for solving nonlinear optimization problems with inequality constraints. *IEEE Transactions on Neural Networks*, 19(8):1340–1353, 2008.

[147] Y. Xia and M. S. Kamel. A cooperative recurrent neural network for solving L1 estimation problems with general linear constraints. *Neural Computation*, 20(3):844–872, 2008.

[148] J. Xu, Y. Cao, D. Pi, and Y. Sun. An estimation of the domain of attraction for recurrent neural networks with time-varying delays. *Neurocomputing*, 71(7-9):1566–1577, 2008.

[149] R. Xu, G. K. Venayagamoorthy, and D. C. Wunsch II. Modeling of gene regulatory networks with hybrid differential evolution and particle swarm optimization. *Neural Networks*, 20(8):917–927, 2007.

[150] Y. Yamashita, M. Takahasi, T. Okumura, M. Ikebuchi, H. Yamada, M. Suzuki, K. Okanoya, and J. Tani. Developmental learning of complex syntactical song in the Bengalese finch: A neural network model. *Neural Networks*, 21(9):1224–1231, 2008.

[151] Y. Yamashita and J. Tani. Emergence of functional hierarchy in a multiple timescale neural network model: a humanoid robot experiment. *PLoS Computational Biology*, 4(11), 2008.

[152] T. Yamazaki and S. Tanaka. The cerebellum as a liquid state machine. *Neural Networks, Special Issue on Echo State and Liquid State networks*, 20(3):290–297, 2007.

[153] W. Yilei, S. Qing, and L. Sheng. A normalized adaptive training of recurrent neural networks with augmented error gradient. *IEEE Transactions on Neural Networks*, 19(2):351–356, 2008.

[154] S. J. Yoo, J. B. Park, and Y. H. Choi. Adaptive output feedback control of flexible-joint robots using neural networks: Dynamic surface design approach. *IEEE Transactions on Neural Networks*, 19(10):1712–1726, 2008.

[155] W. Yu. Multiple recurrent neural networks for stable adaptive control. *Neurocomputing*, 70(1-3):430–444, 2006.

[156] W. Yu, J. Cao, and G. Chen. Stability and Hopf bifurcation of a general delayed recurrent neural network. *IEEE Transactions on Neural Networks*, 19(5):845–854, 2008.

[157] Z. Zeng and J. Wang. Global exponential stability of recurrent neural networks with time-varying delays in the presence of strong external stimuli. *Neural Networks*, 19(10):1528–1537, 2006.

[158] Z. Zeng and J. Wang. Improved conditions for global exponential stability of recurrent neural networks with time-varying delays. *IEEE Transactions on Neural Networks*, 17(3):623–635, 2006.

[159] Z. Zeng and J. Wang. Analysis and design of associative memories based on recurrent neural networks with linear saturation activation functions and time-varying delays. *Neural Computation*, 19(8):2149–2182, 2007.

[160] B. Zhang, S. Xu, and Y. Zou. Improved delay-dependent exponential stability criteria for discrete-time recurrent neural networks with time-varying delays. *Neurocomputing*, 72(1-3):321–330, 2008.

[161] H. Zhang, Z. Wang, and D. Liu. Global asymptotic stability of recurrent neural networks with multiple time-varying delays. *IEEE Trans. Neural Networks*, 19(5):855–873, 2008.

[162] L. Zhang, Z. Yi, and J. Yu. Multiperiodicity and attractivity of delayed recurrent neural networks with unsaturating piecewise linear transfer functions. *IEEE Transactions on Neural Networks*, 19(1):158–167, 2008.