Title: Non-gradient approaches to training recurrent neural networks.
Presenter: Stefan C. Kremer

This presentation will focus on non-gradient training methods for recurrent networks.

Since the discovery of the shrinking gradients problem (Hochreiter, 1991; Bengio et al. 1994), there has been much interest in developing training algorithms that do not suffer from this limitation. The long short-term memory (LSTM) approach (Hochreiter and Schmidhuber, 1997), approaches the problem by using gating units that linearize the activation functions of the units in the network during the training process thus preventing the gradients from deteriorating. In this talk, I will present a number of new methods that use non-gradient techniques to adapt the weights of a recurrent network. The techniques presented will include unsupervised approaches as well as a number of approaches based on simulated annealing.

We begin our discussion with an analysis of a number of unsupervised learning algorithms based on a non-linear generative framework. This framework allows us to easily compare and contrast a variety of different approaches in terms of both representational and learning issues (Barreto, Araijo and Kremer, 2003).

Next we examine the semi-supervised learning paradigm where unlabelled data is used to contribute to a supervised learning process. In particular we describe the utility of this approach in problems where labeled data is limited including bioinformatics.

Then we examine some new work applying simulated annealing to the task of training recurrent networks. While simulated annealing in non-recurrent networks has not been extensively used, because of its comparatively slow convergence relative to gradient based approaches, it more than holds its own in the recurrent domain where it actually outperforms more popular techniques (Unnikrishnan and Venugopal, 1994; Bia, 2001; Sottas and Gerstner, 2002).

Bibliography

Barreto, G., Araijo, A., and Kremer, S.C. (2003) “A Taxonomy for Spatio-temporal Connectionist Networks Revisited: The Unsupervised Case”, *Neural Computation*, **15**(6): pp.1255-1320.

Bengio, Y., Simard, P., and Frasconi, P. (1994) “Learning Long-Term Dependencies with Gradient Descent is Difficult”, *IEEE TNN*, **5**(2), pp. 157-166.

Bia, A. (2001) “Alopex-B: A New, Simpler, Yet Faster Version of the Alopex Traing Algorithm, *International Journal of Neural Systems*, (special issue on “Non-gradient learning techniques”, de Carvalo, A and Kremer, S.C. (eds.)).

Hochreiter, J. (1991) “Untersuchung zu dynamischen neuronalen Netzen”, Intitut fuer Infomatik, Lehrstuhl Prof. Brauer, Technishe Universitaet Muenchen.

Hochreiter, S., and Schmidhuber, J. (1997) “Long Short Term Memory”, *Neural Computation*, **9**, pp. 1681-1726.

Sottas, P. and Gerstner, W. (2002) “Dynamic noise annealing for learning temporal sequences with recurrent neural networks”. In J. R. Dorronsoro (ed.), *ICANN 2002 Proceedings*, pp. 1144-1149.

Unnikrishnan, K. and Venugopal, K. (1994) “Alopex: A Correlation-Based Learning Algorithm for Feedforward and Recurrent Neural Networks”, *Neural Computation*, **6**(3), pp. 469-490.