Ultrafast Neural Network Learning from Uncertain Data

Jacob Barhen & Vladimir Protopopescu

Center for Engineering Science Advanced Research

Computer Science and Mathematics Division

Oak Ridge National Laboratory
Oak Ridge, TN 37831-6355

WSES'2001 International Conference on Neural Networks
February 11-15, 2001 Tenerife, Canary Islands

This research was performed at the Oak Ridge National Laboratory (ORNL), managed by UT-Battelle, LLC for the United States Department of Energy (U.S.DOE) under Contract No. DE-AC05-00OR221725. Funding for the effort was provided by the Basic Energy Sciences Engineering Research Program of the Office of Science, U.S. DOE.

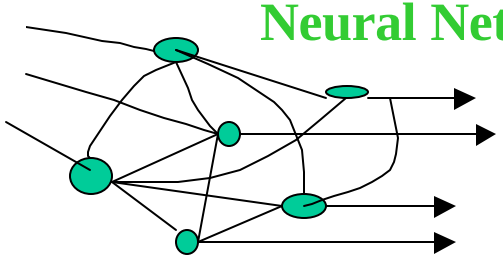
Outline

- Develop **faster learning algorithms** for neural nets
- Develop methodology to **determine the confidence limits of results** predicted by neural network models
- Obtain **best estimates** for network model parameters
  - consistently combine experimental data (sensor measurements) with model output
- Significantly **reduce** the **uncertainties** underlying decision processes based on learning
- Apply new methodology to characterization of oil fields

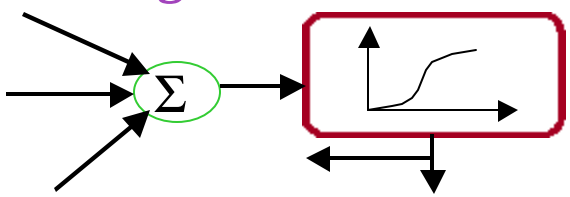
Artificial Neuromorphic Systems

- Large scale nonlinear dynamical systems composed of simple interacting units
- Can carry out useful information processing by means of their collective [emergent] computational properties
- Biological neural networks provide “proof-of-concept” guidelines

Neural Network



Processing Element



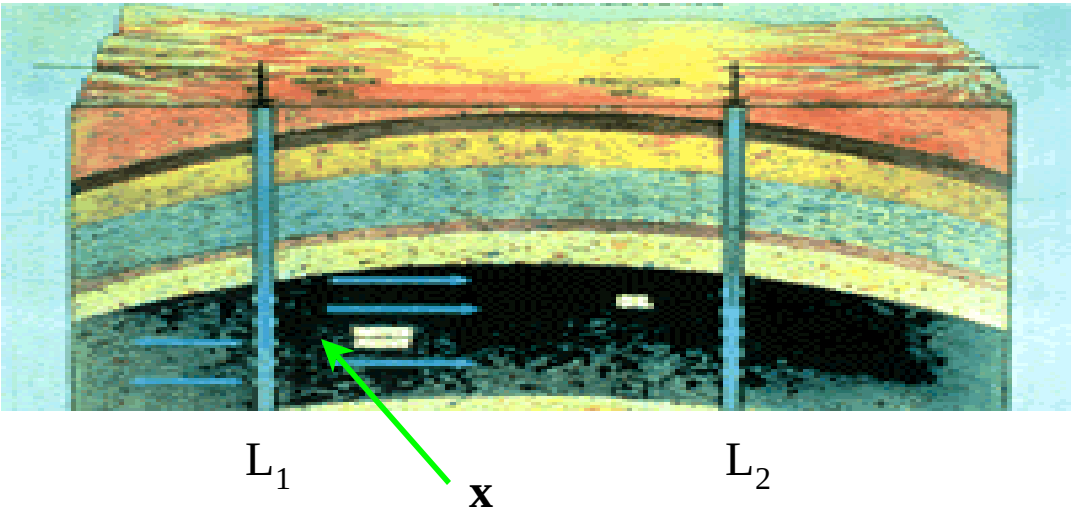
$$C_i \, du_i / dt = \sum W_{ij} \, g(u_j) - u_i / R_i + I_i$$

Artificial Neuromorphic Systems

Learning

Neural nets provide efficient solutions to challenging problems in pattern recognition, signal analysis, and real-time control of complex systems.

At the heart of such advances lies the development of very efficient methodologies for learning



Challenges

- systematic treatment of **uncertainty**
- very **slow learning** for systems of realistic size
- convergence to **local minima** of error function

Oil Field Characterization

- **Goal: given** detailed log data from L_1 and L_2 and seismic survey data between wells, **estimate** subsurface properties at **x**
- But, how much confidence can one place in such network predictions obtained from sensor data corrupted by uncertainty ?

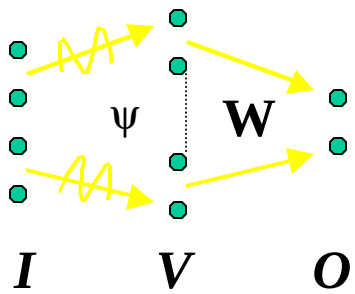
Learning

Development of neural learning algorithms is generally based upon the minimization of an energy–like neuromorphic error function

- pioneering contributions of Grossberg, Werbos, Hopfield
- **gradient-based techniques** have provided the main computational mechanism for the optimization process
  - excessive training times for large networks
- considerable efforts have been devoted by many groups to
  - speed up rate of convergence
    e.g., Barhen & Zak, *IEEE Computer*, 22, 67-76 (1989)
  - compute more efficiently gradients
    e.g., Barhen & Toomarian, *Appl. Math. Lett.*, **3**, 13-18 (1990)
    e.g., Toomarian & Barhen, *U.S Patent No.* 5,930,781 (July 1999)
- novel approaches based on **linear algebraic methods** pioneered by Biegler-König and Bärman
  e.g., Barhen & Protopopescu, *Neural Process. Lett.*, **11**, 113-129 (2000)

DeepNet

ARCHITECTURE



Feedforward flow

- I nodes input layer
- V nodes virtual layer
- O nodes output layer

K cluster patterns

- **Inputs**: Ω_{KI}
- **Outputs**: $\mathbf{R}_{KO}$

Synaptic interconnections

- $\mathbf{W}_{VO}$ linear
- ψ nonlinear

Approach

Training Data

- given L training patterns $\{\Omega_{LI}, \mathbf{R}_{LO}\}$
  - define K clusters: $I \perp K \perp L$: $\{\Omega_{KI}, \mathbf{R}_{KO}\}$
- compute nominal uncertainties for each cluster

Computational Model

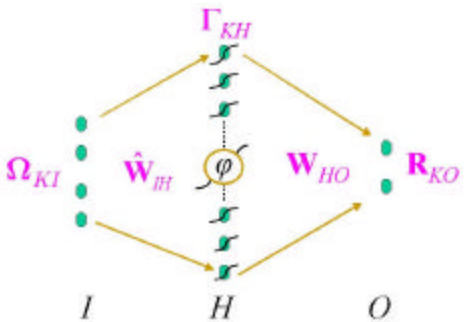
$$\mathbf{R}_{KO} = \mathcal{M}(\Omega_{KI}, \mathbf{W}_{VO})$$

Unique feature

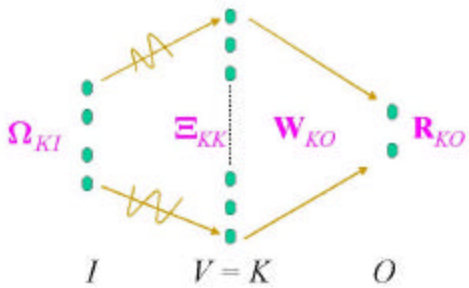
- virtual layer connected to input layer via non-linear ψ transform (**nonlinear weights**)
- ψ maps Ω_{KI} into nonsingular matrix Ξ_{KK}

$$\Xi_{KK} = \psi(\Omega_{KI})$$

DeepNet



$\mathbf{T}_{KO} = \rho^{-1}(\mathbf{R}_{KO})$



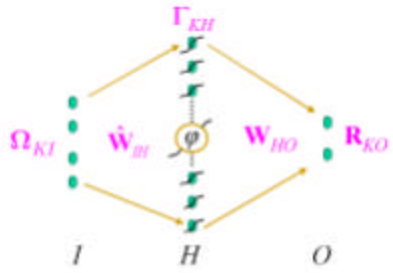
BKB Network

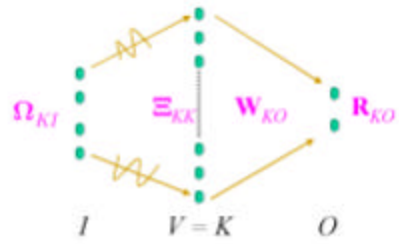
- $\mathbf{P}_{KO} = _{KH}\, \mathbf{W}_{HO}$
- the learning algorithm minimizes the norm $\psi \mathbf{T}_{KO} - \mathbf{P}_{KO}\, \psi$
- since $_{KH}$ **not known**, one sweeps backwards, using initial $\mathbf{W}_{HO}$ guess
  $_{KH}\, \mathbf{W}_{HO} = \mathbf{T}_{KO} = _{KH}\, \mathcal{R}_{HH}\, \mathcal{R}^{-1}_{HH}\, \mathbf{W}_{HO}$
- renormalization matrix $\mathcal{R}_{HH}$ introduced to satisfy sigmoid constraints at HL

DeepNet

- $\mathbf{P}_{KO} = \Xi_{KK}\, \mathbf{W}_{KO}$
- the learning algorithm minimizes the norm $\psi \mathbf{T}_{KO} - \mathbf{P}_{KO}\, \psi$
- solves the system
  $\mathbf{T}_{KO} = \Xi_{KK}\, \mathbf{W}_{KO}$
- since $\mathbf{T}_{KO}$ and Ξ_{KK} are known, we can compute $\mathbf{W}_{KO}$ by SVD of Ξ_{KK}

DeepNet





BKB Network

- solve for $\hat{\mathbf{W}}_{IH}$ the system

$$\Omega_{KI}\,\hat{\mathbf{W}}_{IH} = j^{-1}(\Gamma_{KH}\,\mathcal{R}_{HH})$$

- inversion of Ω_{KI} introduces an unavoidable loss of accuracy, since $K \pi I$

- forward sweep yields $\mathbf{P}_{KO}$

$$\mathbf{P}_{KO} = j\,(\Omega_{KI}\,\hat{\mathbf{W}}_{IH})\,\mathcal{R}_{HH}^{-1}\,\mathbf{W}_{HO}$$

DeepNet

- in *virtual* layer architecture the matrix Ξ is nonsingular

$$\mathrm{X}_{KK}^{(i)}(k,l) = 1 - \frac{\left|\omega_k^{(i)} - \omega_l^{(i)}\right|}{D^{(i)}}$$

$$\det(\mathrm{X}_{KK}) = \prod_{i=1}^{I}\left(\frac{1}{2}\prod_{k=1}^{K-1}\frac{2u^{(i)}(k)}{D^{(i)}}\right) > 0$$

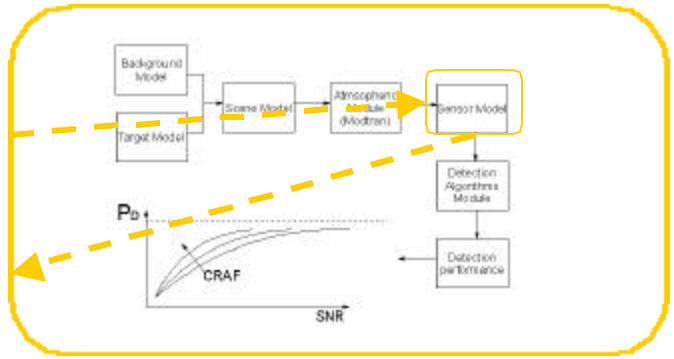
➡ new ***DeepNet*** methodology will enable **accurate ultrafast** (***single iteration***) **learning**

Learning under Uncertainty

Recognized need for **information fusion** tools that explicitly account for **data and model uncertainties**. The design of complex multisensor-based target–detection / tracking architectures *illustrates* typical application.

For each model

- inputs
- parameters
- system responses
  i.e., *outputs*



The methodology we are developing has tree primary goals:

- determine ***confidence limits*** of ***neural net – predicted*** results
- consistently ***combine*** sensor measurements with computational results
  - obtain ***best estimates*** of model parameters
  - ***reduce uncertainties*** in estimates
- apply to CESAR tasks: **robotics**, **seismic analysis**, **quantum dots**

Critical Requirements
for Learning under Uncertainty

Techniques based upon model sensitivities enable **robust learning and generalization by a neural network** via new information fusion

Such methods should incorporate five **key capabilities**

- guarantee that no ***important effects*** are overlooked
  - ➡ *full set of **sensitivities***
- fast processing of large data sets
  - ➡ *efficient computation of sensitivities*
- systematic treatment of ***nonlinearities***
- inclusion, where relevant, of ***time dependence*** *inputs, model parameters, responses*
- comprehensive method for **combining experimental** (i.e., **sensor**) **data** and **model results**; goal is to ***reduce the uncertainties***

J. Barhen and V. Protopopescu, in *DARS*, **4**, 403-413, Springer (2000)

F. Aminzadeh and J. Barhen, *Petrol. Sci. & Eng.*, **24**, 49-56 (2000)

F. Aminzadeh and J. Barhen, *Comp. & Geosci.*, **26**, 869-875 (2000)

NOGA

Sensitivities

- provide a systematic way to **propagate uncertainties** in complex, non stationary, nonlinear models

$$S^{\acute{\imath}\grave{\imath}}_{ni} = \partial q^{\acute{\imath}}_{n} / \partial a^{\grave{\imath}}_{i}$$

system parameters: $\mathbf{a} = \{\mathbf{W}, \quad \}$

network computed responses: $\mathbf{q} = \rho(\mathbf{p})$

experimental response: $\mathbf{r}$

- determine and **rank** the **importance** of parameters and input data to computed quantities of interest (*system responses*)
- computed using **adjoint operator** methods or automatic differentiation codes for recurrent nets
- nominal ***uncertainties*** quantified in terms of **covariance matrices** $C_{\mathbf{aa}} = \left\langle \Delta\mathbf{a}\,\Delta\mathbf{a}^{t} \right\rangle$

Bayesian Inference

- **information fusion goal** is to combine consistently ***computational results and measurements***
- ***minimize a generalized Bayesian loss function with constraints***

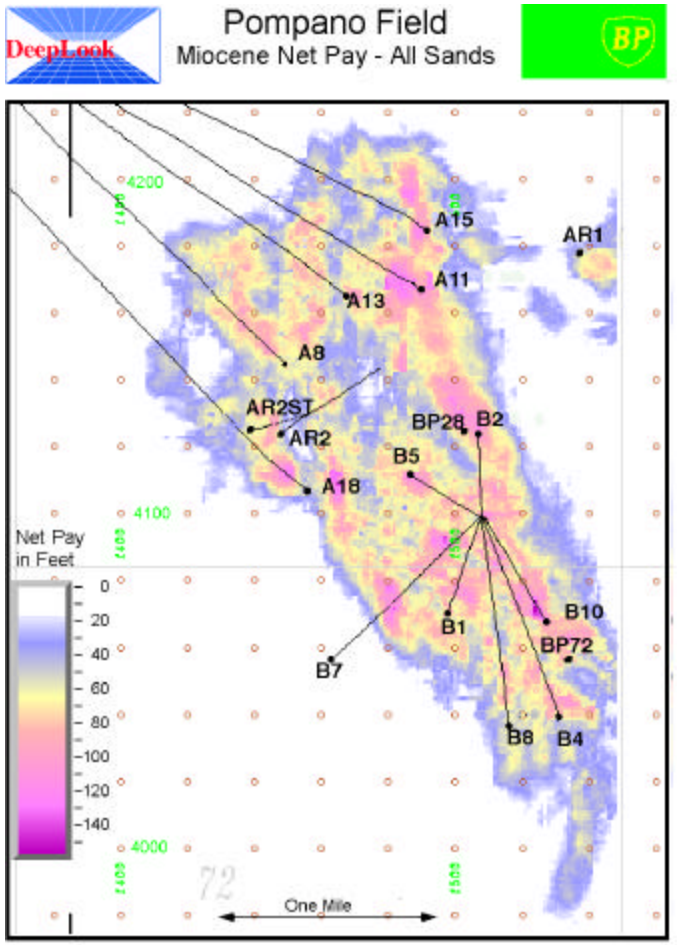
$$L = (\hat{\mathbf{q}}-\mathbf{r} \,|\, \hat{\mathbf{a}}-\mathbf{a} \,|\, \ast)\begin{pmatrix} C_{\mathbf{rr}} & C_{\mathbf{ra}} & \blacktriangle \\ C_{\mathbf{ar}} & C_{\mathbf{aa}} & \blacktriangledown \\ \blacktriangle & \blacktriangledown & \blacklozenge \end{pmatrix}^{-1}\begin{pmatrix} \hat{\mathbf{q}}-\mathbf{r} \\ \hat{\mathbf{a}}-\mathbf{a} \\ \bullet \end{pmatrix} + \mathbf{l}^{T}\mathbf{c}[\hat{\mathbf{a}},\mathbf{a},\hat{\mathbf{q}},\mathbf{q},\mathbf{r},\ldots,\mathbf{S},\ldots]$$

- optimization uses the inverse of the total covariance matrix as the **natural metric** of the calculational manifold
- solution of optimality equations yields **best estimates** for parameters and **reduced uncertainties**

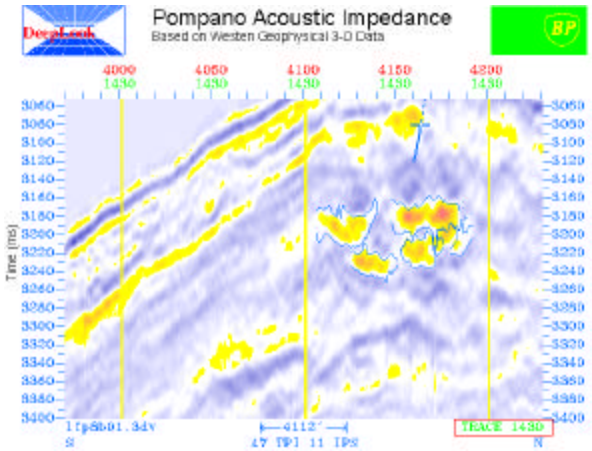
Oil Field Characterization

Framework

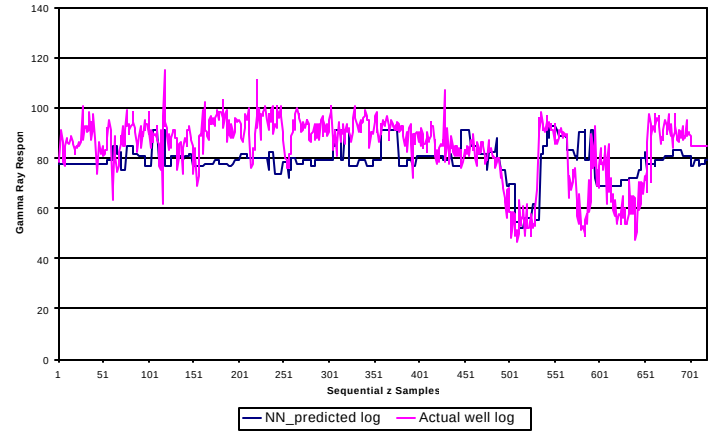
- Ability to predict location of remaining oil in neighborhood of existing production wells is of vital economic importance to petroleum industry
- **Neural nets** are used to capture the relationship between seismic survey data and well logs
- Avoid costly drilling of new wells
- Demonstrate capability of ***DeepNet*** to capture this extremely complex nonlinear relationship

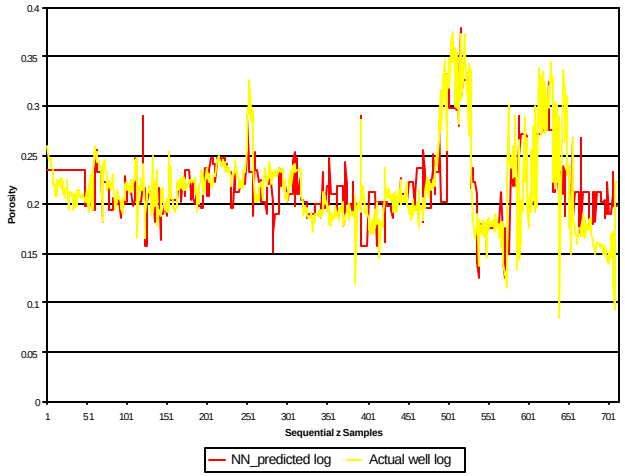


DeepNet Results

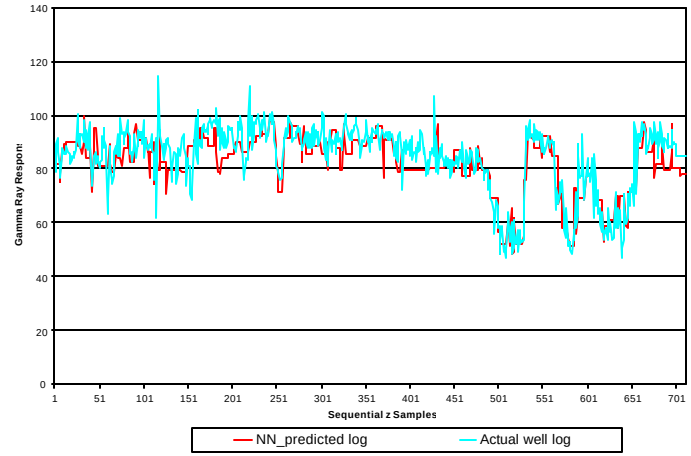


Prediction of best-estimates: **gamma ray** log. Current state-of-art method (Nadaraya-Watson by Rao) Data: B10_5S (well B10)





DeepNet (BKB) prediction of best-estimates **porosity** log. Data: B10_5S (well B10)



DeepNet (BKB) prediction of best-estimates **gamma ray** log. Data: B10_5S (well B10)

Summary

We have developed new methods for neural net learning under uncertainty

- **ultrafast** (single iteration) ***DeepNet*** neural net learning code provides the computational framework for modeling
- **information fusion** module ***NOGA*** uses ***sensitivity matrices*** in conjunction with minimization of a constrained generalized Bayesian loss function to ***combine model prediction and experimental results***
- best-estimates for model parameters and ***reduced uncertainties***

DeepNet / ***NOGA*** expected to provide **enabling capabilities** for several CESAR basic research activities

- pseudo-logs prediction in the neighborhood of petroleum reservoir wells
- tracking in cooperating robots
- pattern classification using quantum dot arrays
- synchronization of arrays of lasers