Time series prediction

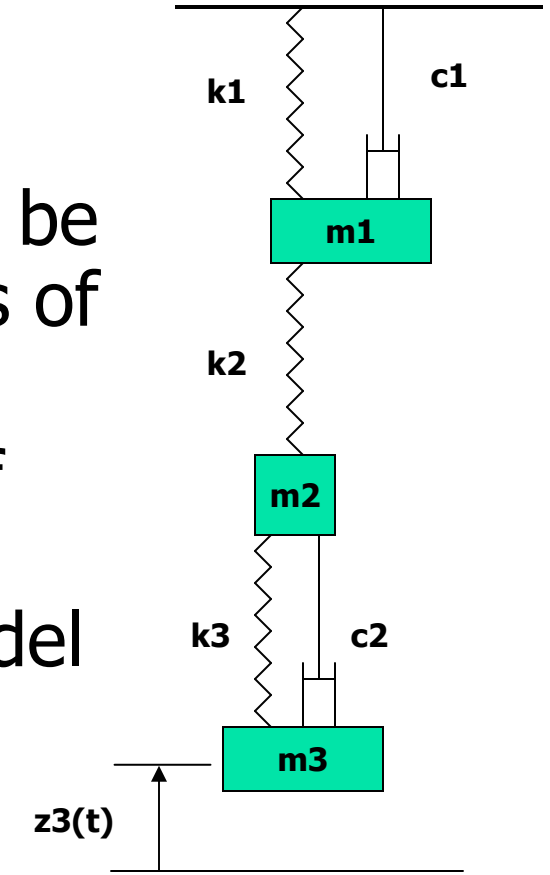
- Given measurements of a time series x(k) at time points k,k-1,k-2,k-3,…,k-n, find an estimate $\hat{x}(k+1)$ of x(k+1).

- Applications in macroeconomics, finance, meteorology, geology, health care etc.

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Model-based prediction

- In many cases, e.g. mechanical systems, a model of the system can often be generated. All that need to be determined then are the parameters of the model.
- In other cases, e.g. for prediction of weather, financial data etc., it is generally difficult to find a good model of the system => use e.g. ANN.



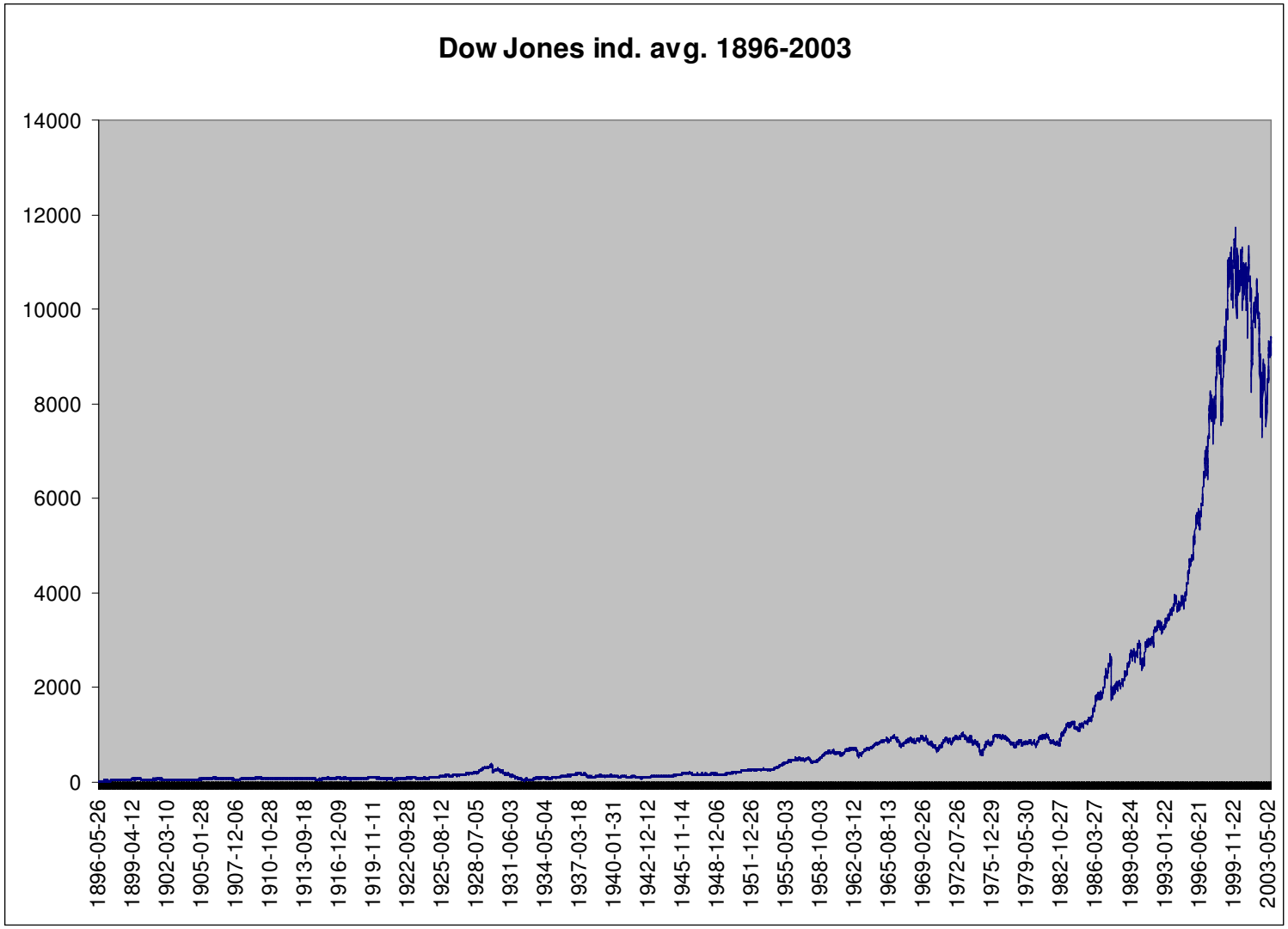
Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Data set example

DJIA1971_1974
4 1

0.543400	0.576571	0.538600	0.536514	0.452657
0.576571	0.538600	0.536514	0.452657	0.430314
0.538600	0.536514	0.452657	0.430314	0.445771
0.536514	0.452657	0.430314	0.445771	0.516886
0.452657	0.430314	0.445771	0.516886	0.594714
0.430314	0.445771	0.516886	0.594714	0.607857
0.445771	0.516886	0.594714	0.607857	0.602857
0.516886	0.594714	0.607857	0.602857	0.594914
0.594714	0.607857	0.602857	0.594914	0.540886
0.607857	0.602857	0.594914	0.540886	0.554229
0.602857	0.594914	0.540886	0.554229	0.554029
0.594914	0.540886	0.554229	0.554029	0.499571
0.540886	0.554229	0.554029	0.499571	0.435343

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

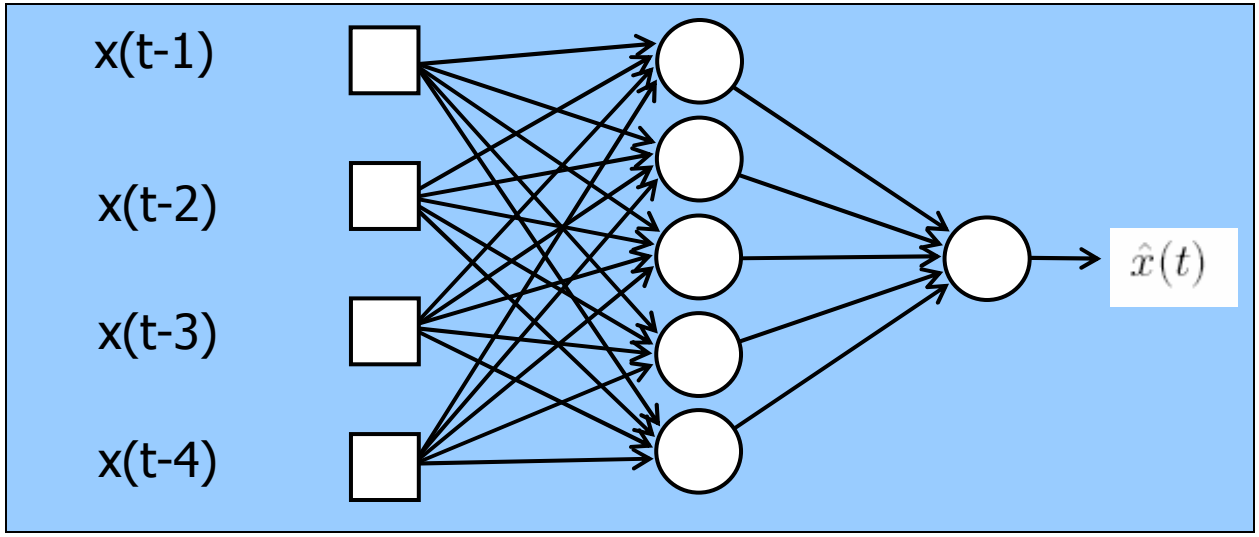


Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Using FFNN for prediction

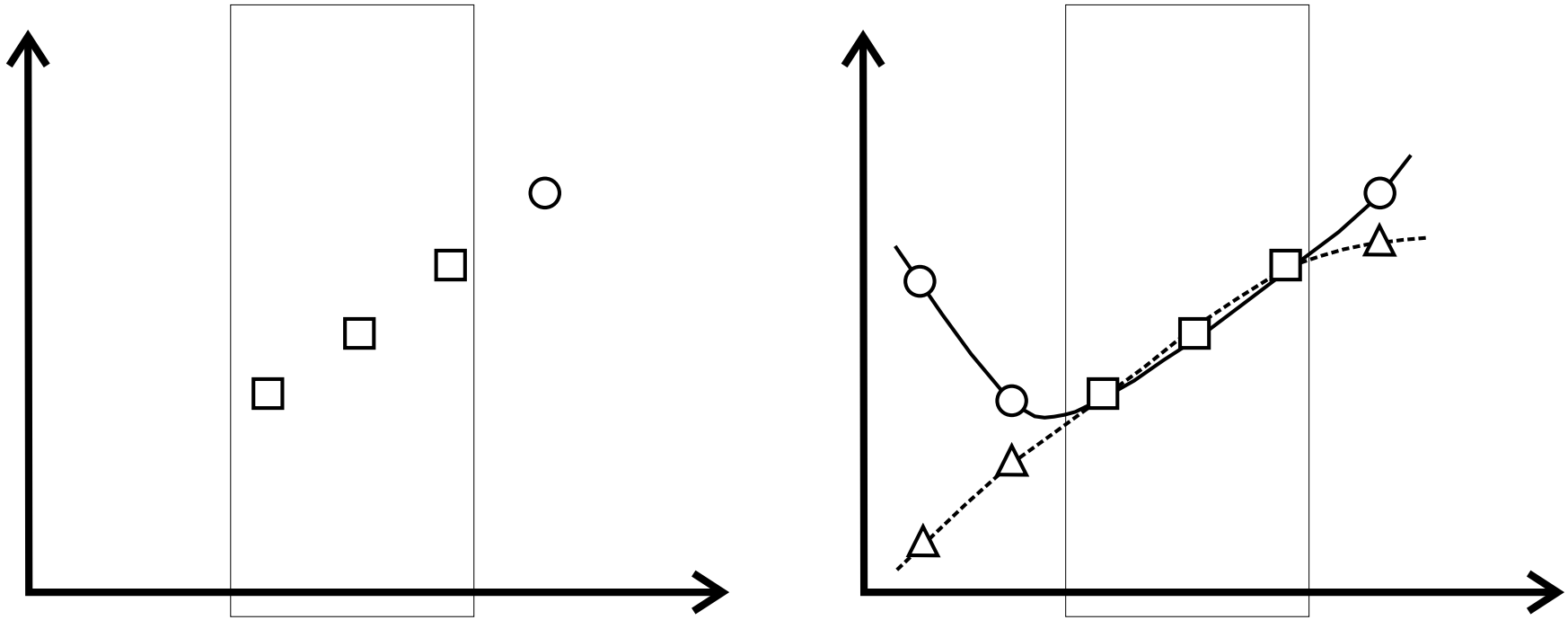
- Common method:

  FFNN + backpropagation



Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Problem: no dynamic memory



Also: fixed network structure if backpropagation is used – the chosen structure can turn out not to be optimal.

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Using RNN for prediction

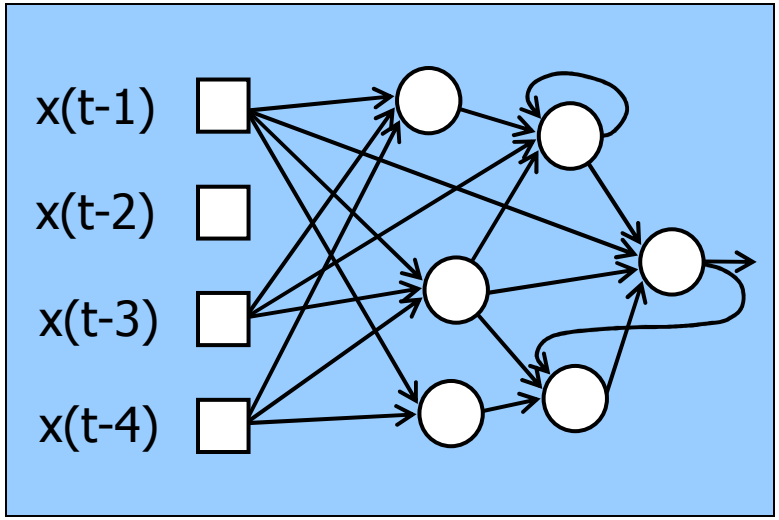
- Both problems can be solved if instead an RNN is used.

- Additional problems:

  (1) Signal flow through the RNN

  (2) Setting the structure

  (3) Training

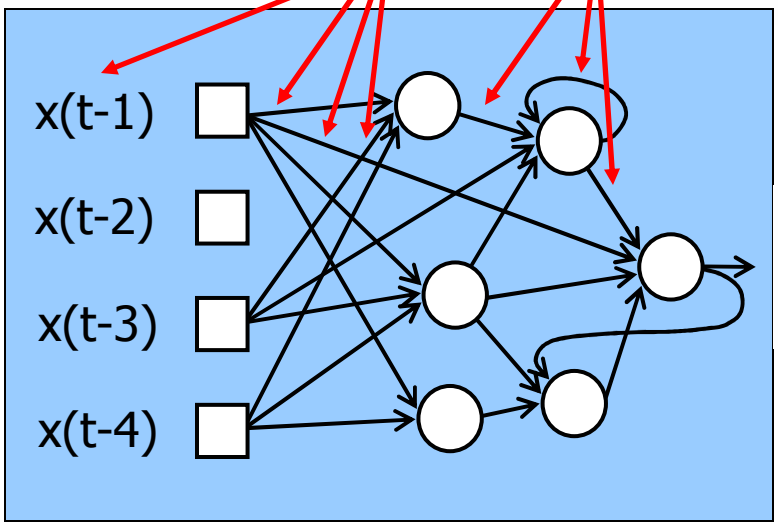


Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Using RNN for prediction

- Signal flow: use continuous-time dynamics

$$\tau_i \dot{y}_i + y_i = \sigma\left(b_i + \sum_j w_{ij}^{\mathrm{IN}} I_j + \sum_j w_{ij} y_j\right), \quad i = 1, \ldots, n$$

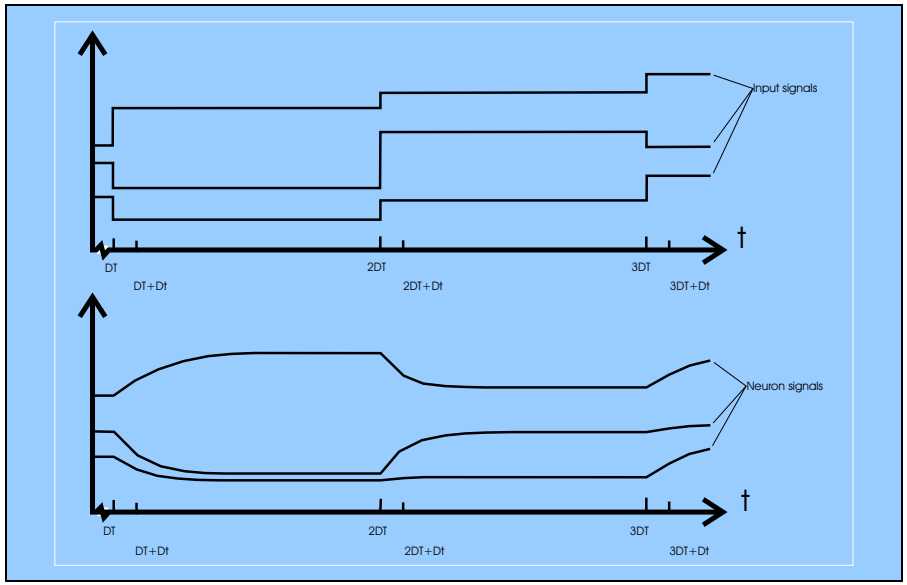


$$e_{\mathrm{RMS}} = \sqrt{\frac{1}{N-L+1}\sum_{j=L}^{N}(\hat{X}(j) - X(j))^2}$$

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Using RNN for prediction

- The continuous-time dynamics unfolds (within the RNN) between successive inputs



Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

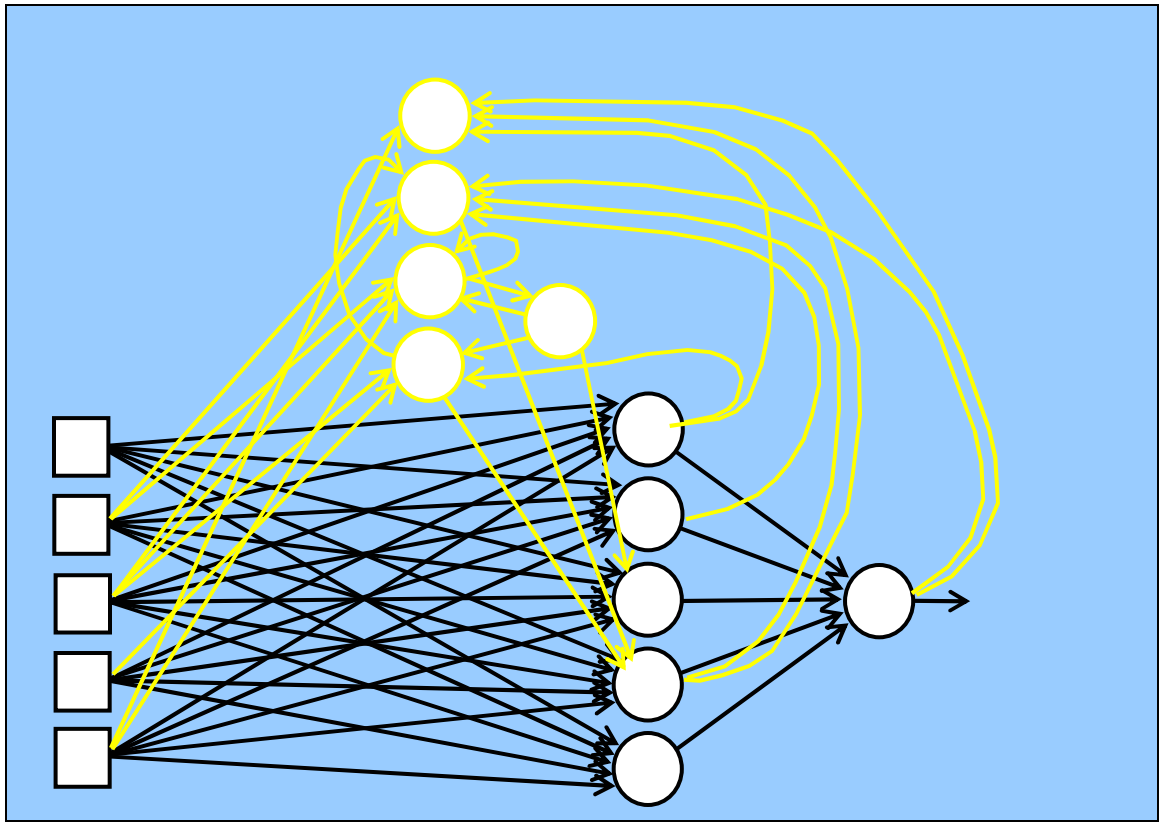
Using EA for generating RNN

- Problem: slow to evolve from scratch (if arbitrary network structure is allowed)

- Solution: First generate an FFNN using backpropagation, *then* apply an EA (starting from the FFNN), to generate an RNN.

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Illustration of the procedure



Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Methods (summary)

(1) FFNN + Backpropagation

(2) EA + RNN (from random initiation)

(3) EA + RNN (from FFNN, generated using backpropagation)

(4) Naive predictor: $\hat{x}(t+1) = x(t)$

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Time series types

- Original series
- Difference series $x_{\mathrm{d}}(t) = x(t+1) - x(t)$
- Relative difference series $x_{\mathrm{d}}(t) = \dfrac{x(t+1) - x(t)}{x(t)}$

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Results (best predictors)

Series	FFNN	RNN (from random)	RNN (from FFNN)
USD-JPY	2.20%	1.93%	3.12%
USD-JPY (d)	2.42%	4.39%	4.01%
US unempl.	1.19%	5.50%	4.68%
US unempl. (d.)	2.04%	4.09%	4.12%
DJIA (r.d.)	5.08%	6.01%	6.00%

(All results shown were obtained on the *validation* data. The percentages indicate the improvements relative to a naive predictor, $\hat{x}(t+1) = x(t)$)

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Results: average and st.dev.

Series	FFNN	RNN (from random)	RNN (from FFNN)
USD-JPY	1.38% 1.15%	-2.15% 6.43%	2.37% 0.62%
USD-JPY (d)	1.68% 0.50%	2.64% 1.52%	3.44% 0.30%
US unempl.	-2.37% 2.21%	2.18% 1.84%	3.07% 0.71 %
US unempl. (d.)	0.44% 0.94%	3.33% 0.51%	3.74% 0.31%
DJIA (r.d.)	4.37% 0.47%	5.09% 0.50%	5.72% 0.16%

(Improvements are measured relative to a naive predictor, $\hat{x}(t+1) = x(t)$)

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

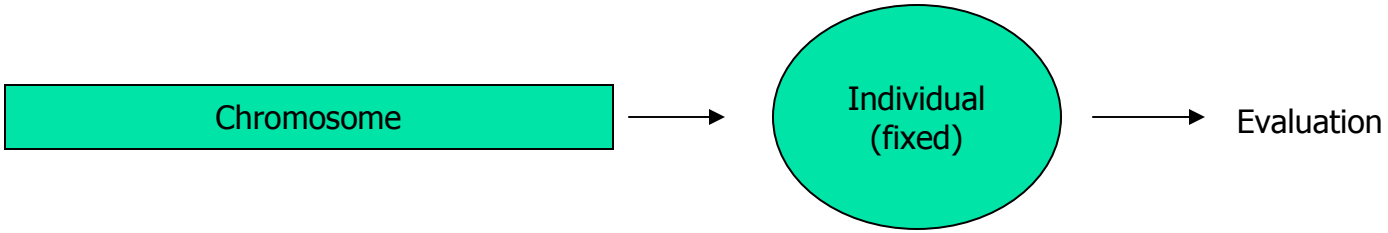
Conclusions

- RNNs generated by EAs (either starting from a random initial population or an FFNN) consistently outperform FFNNs and naive predictors. Thus short-term memory appears to be crucial.
- The improvements compared to the naive predictor are rather small, but significant (and consistent).
- The method of first generating an FFNN and then evolving an RNN generates better results on average, an in *much shorter time* than evolution from a random initial population.

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Evolution of learning rules in ANN

- Normal GA operation: fixed individual decoded from genome

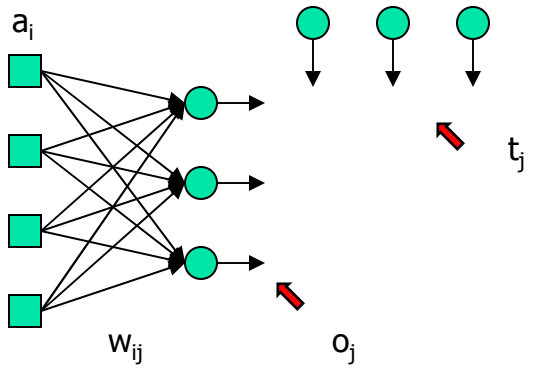


- By contrast, biological organisms (even simple ones) are capable of *learning*.

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Evolution of learning rules in ANN

- An example of the evolution of learning: Chalmers (1990) http://citeseer.ist.psu.edu/chalmers90evolution.html
- Chalmers evolved learning rules for feedforward, single-layer neural networks



Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Evolution of learning rules in ANN

- The delta rule:

$$\Delta w_{ij} = \eta(t_i - o_i)a_j$$

- Chalmers' representation:

$$\Delta w_{ij} = k_0(k_1 w_{ij} + k_2 a_j + k_3 o_i + k_4 t_i + k_5 w_{ij} a_j + k_6 w_{ij} o_i + k_7 w_{ij} t_j + k_8 a_j o_i + k_9 a_j t_i + k_{10} o_i t_i)$$

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Evolution of learning rules in ANN

- Encoding scheme:

11011 000 000 000 000 000 000 010 110 000

$k_0 = 2^{j-1}$ if $j > 0$, otherwise 0 (four last bits used for j ([0,15]), the first bit encodes the sign.

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Evolution of learning rules in ANN

- Rules were applied to networks with 2-7 inputs and one output.
- The rules encoded by the chromosomes were applied to a number of different input-output mappings
- The evolved best rules were essentially versions of the delta rule, e.g.

$$\Delta w_{ij} = 8(t_i - o_i)(a_j - 0.5)$$

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Evolution of learning rules in ANN

- More advanced example (multi-layer networks), see Radi and Poli (2002), http://citeseer.ist.psu.edu/radi02discovering.html

$$\Delta w_{ij}^{l}(s) = \begin{cases} F_o(w_{ij}^{l}, o_{jp}^{l}, t_{ip}, o_{ip}^{l+1}) & \text{for the output layer,} \\ F_h(w_{ij}^{l}, o_{jp}^{l}, o_{ip}^{l+1}, \varepsilon_{ip}^{l+1}) & \text{for the hidden layers} \end{cases}$$

- Evolution in two stages: first evolve learning rules for the output layer, using standard backprop (SBP) for the hidden layer, and then evolving the rules for the hidden layer

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Evolution of learning rules in ANN

- Modified (compared to SBP) rules were found that outperformed SBP – see the paper for details. Example:

Vowel			
Algorithm	N	% of success	Std. Dev.
SBP	8	32	2.1
SBP	12	39	1.8
SBP	16	45	1.6
SBP	24	49	1.3
SBP	32	50	1.1
$NLR_o + NLR_h$	8	48	1.5
$NLR_o + NLR_h$	12	50	1.2
$NLR_o + NLR_h$	16	58	1.1
$NLR_o + NLR_h$	24	63	1.0
$NLR_o + NLR_h$	32	65	1.0

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

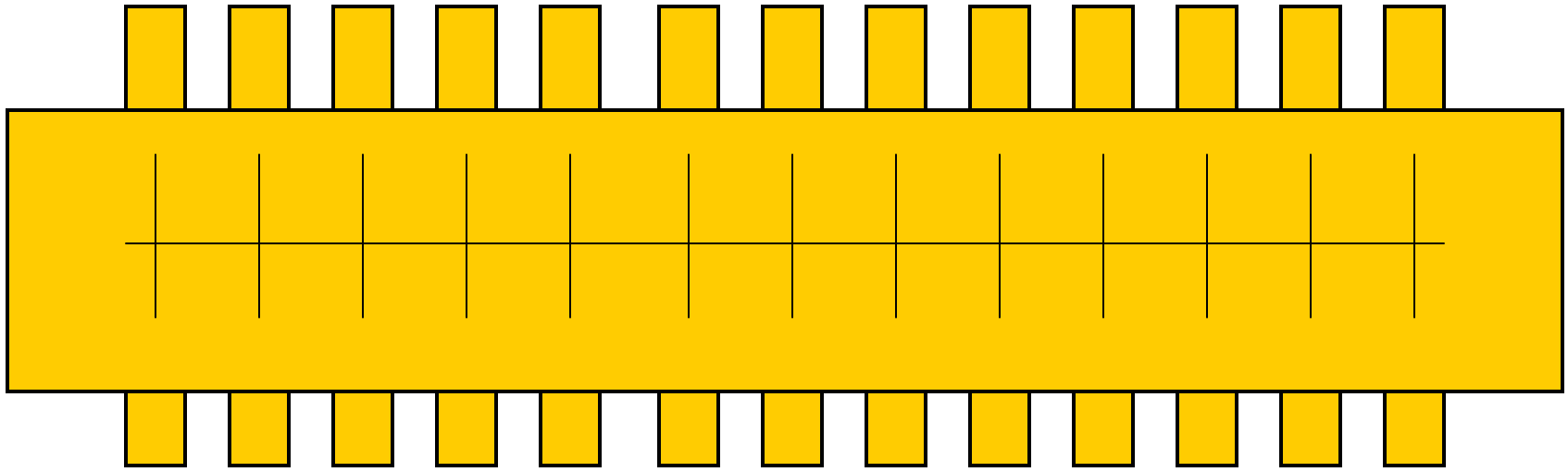
Reconstruction of extinct animals in the computer

- ALIFE VI, 1998
  http://alife6.alife.org/abstracts/RE29.html
- The authors tried to recreate the motion patterns of an extinct (530 Myr) animal (*Anomalocaris*)

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Reconstruction of extinct animals in the computer

- Representation of the body:



Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde

Reconstruction of extinct animals in the computer

- The artificial animal moved in a simplified hydrodynamical model.
- The force in each joint varied according to a trigonometric rule.
- Fitness measure: Motion speed
- Wave-like swimming patterns (similar to those of a stingray) were found to be most efficient

Mattias Wahde, PhD, associate professor, Chalmers University of Technology
e-mail: mattias.wahde@me.chalmers.se www: www.me.chalmers.se/~mwahde