Foundations of Cognitive Dynamic Systems

Simon Haykin
McMaster, University
Hamilton, Ontario, Canada

email: haykin@mcmaster.ca
Web site: http://soma.mcmaster.ca

Outline of The Lecture

1. Introduction
2. A Simplistic View of Cognition
3. Cognitive Dynamic Systems Defined
4. Emerging Applications
5. Foundational Disciplines Involved in Cognitive Dynamic Systems
6. Global Feedback
7. Why Sub-optimality Should be the Objective of Cognitive Dynamic Systems?
8. The Bayesian Filter: A powerful tool for cognitive information processing
9. The Cubature Kalman Filter
10. Supervised Learning

Concluding Remarks

Acknowledgements

1. Introduction
Point-of-View Article, Proc. IEEE
Nov. 2006.

I see the emergence of a new discipline, called Cognitive Dynamic Systems[1], which builds on ideas in statistical signal processing, stochastic control, and information theory, and weaves those well-developed ideas into new ones drawn from neuroscience, statistical learning theory, and game theory. The discipline will provide principled tools for the design and development of a new generation of wireless dynamic systems exemplified by cognitive radio and cognitive radar with efficiency, effectiveness, and robustness as the hallmarks of performance.

---

1. S. Haykin, Cognitive Dynamic Systems, book under preparation.

2. A Simplistic View of Cognition

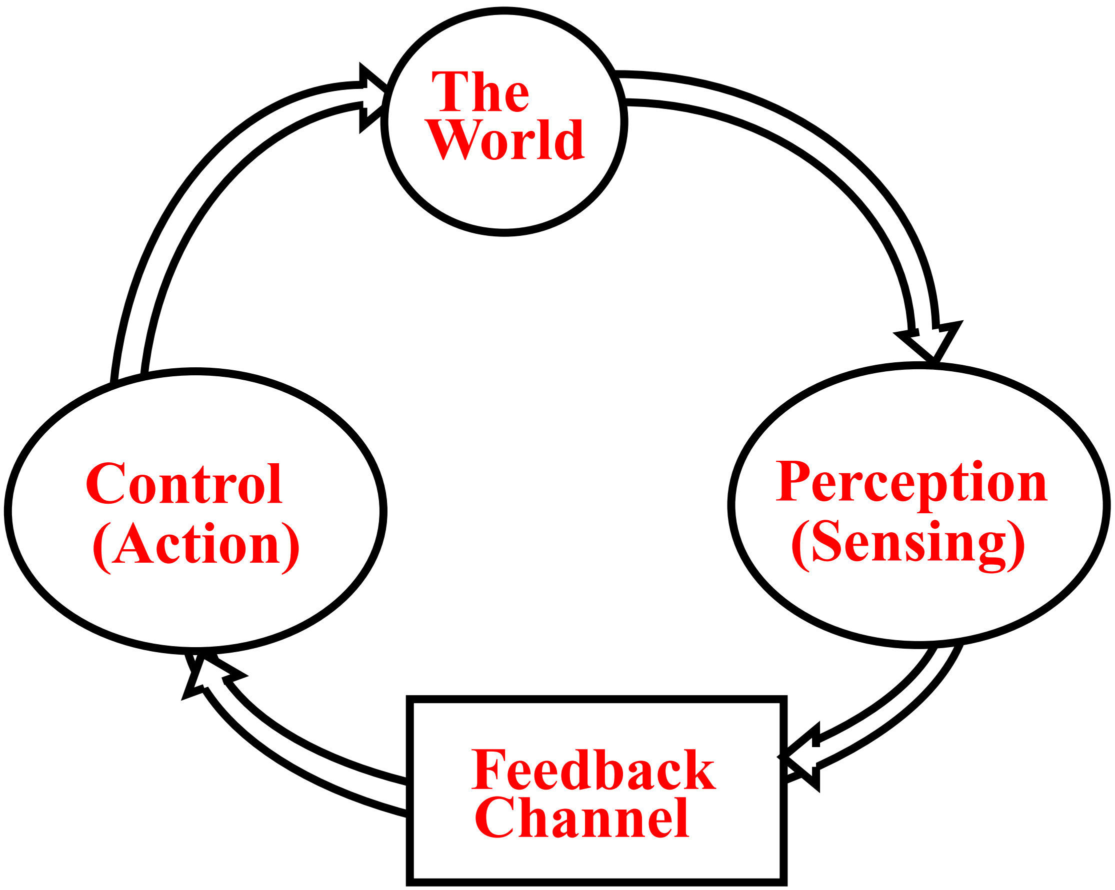

Figure 1. Human Cognitive Cycle in its most basic form

3. Cognitive Dynamic Systems Defined

A Cognitive Dynamic System is a complex system, capable of emergent behaviour.

It processes information over the course of time by performing the following functions:

- ***sense* (perceive) the environment;**
- ***learn* from the environment and adapt to its statistical variations;**
- **build a *predictive model* on prescribed aspects of the environment;**
- **develop *rules of behaviour* so as to act on (control) the environment; and do all of this in real time for the purpose of executing prescribed tasks, in the face of environmental uncertainties, *efficiently and reliably in a cost-effective manner.***

4. Emerging Applications

Cognitive radio

Cognitive radar

Cognitive car

•

•

•

Cognitive Information Processing

Cognitive computation (including software)

Cognitive optimization

Cognitive Radio Networks

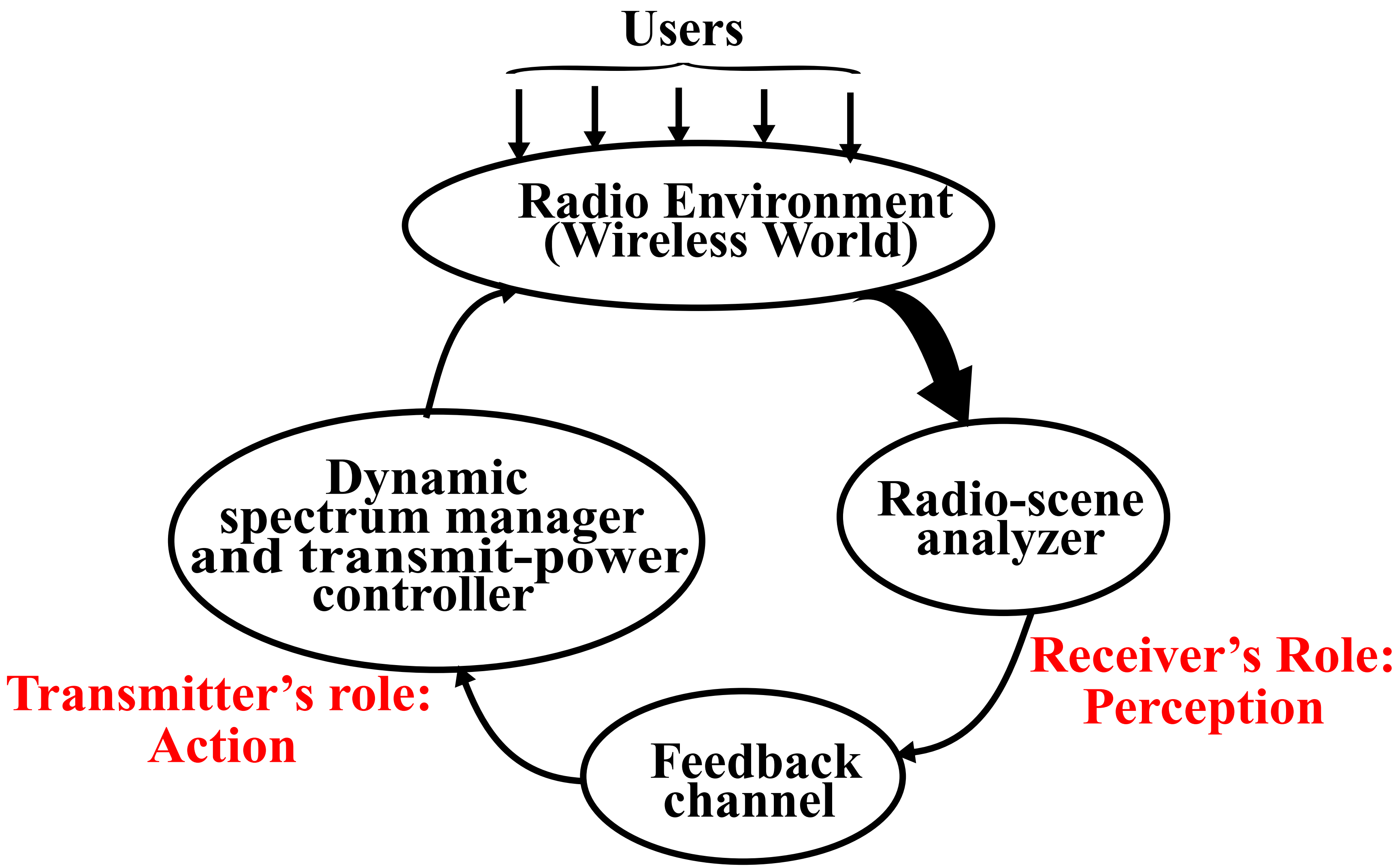

Figure 2. Basic signal-processing cycle, as seen by a single user (transceiver).

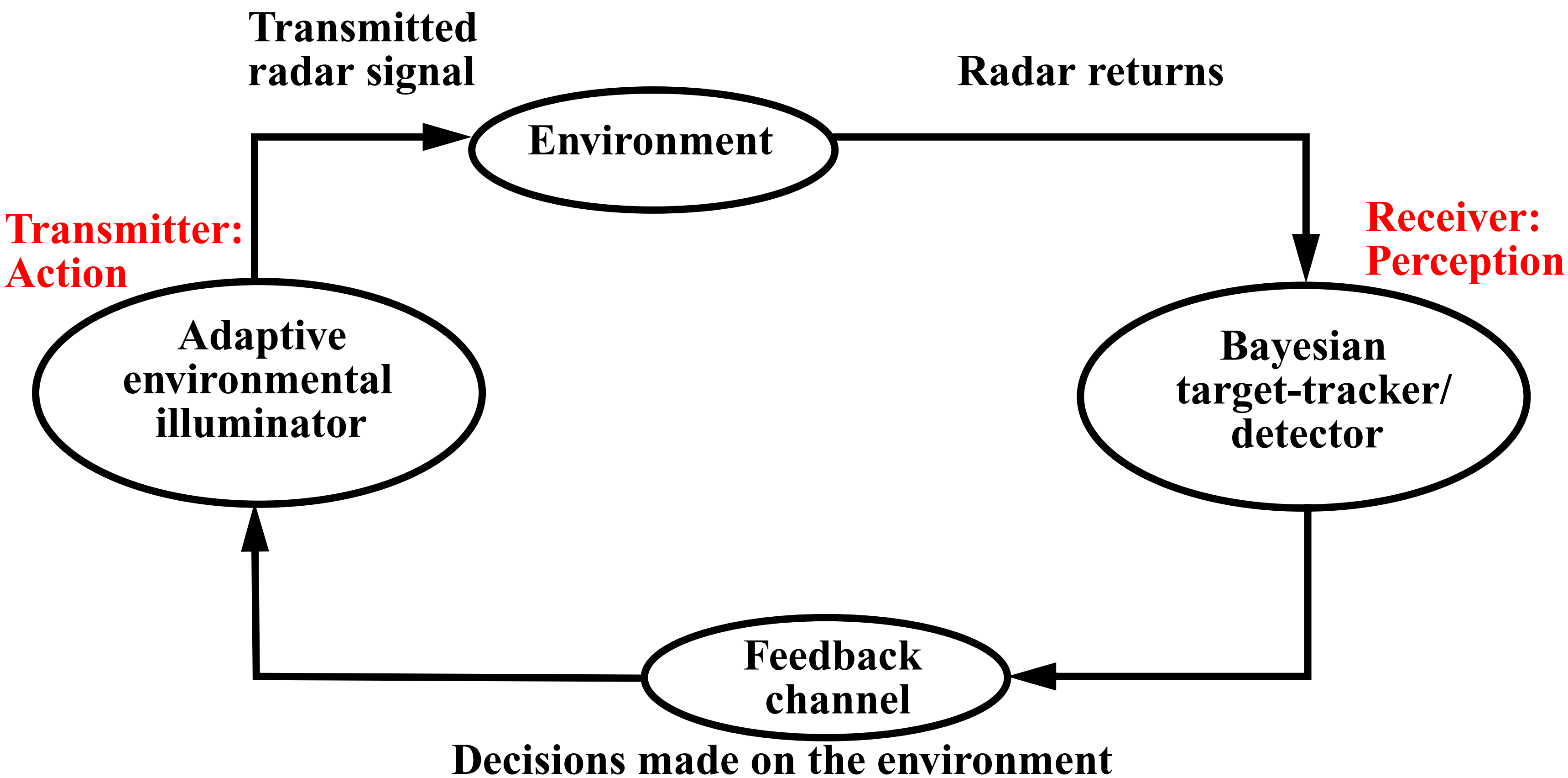

Figure 3. Cognitive tracking radar

5. Foundational Disciplines Involved in Cognitive Dynamic Systems

(i) Bayesian Theory

(ii) Information Theory

(iii) Control Theory:

- Nonlinear filtering
- Dynamic programming

(iv) Learning Theory

(v) Complexity Theory

6. Global Feedback

A Facilitator of Computational Intelligence

- **The human brain is a living example of a cognitive dynamic system with global feedback in many of its parts, be that the visual system, auditory system, or motor control.**
- **Global feedback is responsible for the coordination of different constituents of a cognitive dynamic system.**
- **The emergent behaviour of a cognitive dynamic system is due to the global feedback.**
- **Global feedback is an inherent property of all cognitive dynamic systems, but global feedback by itself will *not* make a dynamic system cognitive.**

7. Why sub-optimality should be the objective of cognitive dynamic systems?

- **Optimality of performance versus robustness of behaviour:**
  A challenge in system design.

- **Global optimality of a cognitive dynamic system is not practically feasible:**

  - **Large-scale nature of the system**
  - **Infeasible computability**
  - **Curse-of-dimensionality**

  Hence, the practical requirement of having to settle for a sub-optimal solution of the system design

- **Trade-off global optimality for computational tractability and robust behaviour.**

Criterion for sub-optimality

DO AS BEST AS YOU CAN, AND NOT MORE

- **This statement is the essence of what the human brain does on a daily basis:**

  Provide the "best" solution in the most reliable fashion for the task at hand, given limited resources.

- **Key question: How do we define "best"?**

8. The Bayesian Filter: A powerful tool for cognitive information processing

Problem statement:

Given a nonlinear dynamic system, estimate the *hidden state* of the system in a recursive manner by processing a sequence of *noisy observations* dependent on the state.

- **The Bayesian filter provides a unifying framework for the optimal solution of this problem, at least in a conceptual sense.**

  Unfortunately, except in a few special cases, the Bayesian filter is not implementable in practice -- hence the need for approximation.

Bayesian Filter (continued)

State-space Model

1. **System (state) Model**

   $$\mathbf{x}_{t+1} = \mathbf{a}(\mathbf{x}_t) + \omega_t$$

2. **Measurement model**

   $$\mathbf{y}_t = \mathbf{b}(\mathbf{x}_t) + \nu_t$$

where t **= discrete time**

$\mathbf{x}_t$ **= state at time** t

$\mathbf{y}_t$ **= observation at time** t

ω_t **= dynamic noise**

ν_t **= measurement noise**

Bayesian Filter (continued)

Assumptions:

- **Nonlinear functions $a(\cdot)$ and $b(\cdot)$ are known**

- **Dynamic noise ω_t and measurement noise ν_t are statistically independent Gaussian processes of zero mean and known covariance matrices.**

Bayesian filter (continued)

Time-update equation:

$$\underbrace{p(\mathbf{x}_t|\mathbf{Y}_{t-1})}_{\text{Predictive distribution}} = \int_{R^n} \underbrace{p(\mathbf{x}_t|\mathbf{x}_{t-1})}_{\text{Prior distribution}} \underbrace{p(\mathbf{x}_{t-1}|\mathbf{Y}_{t-1})}_{\text{Old posterior distribution}} d\mathbf{x}_{t-1}$$

where R^n denotes the n-dimensional state space.

Measurement-update equation:

$$\underbrace{p(\mathbf{x}_t|\mathbf{Y}_t)}_{\text{Updated posterior distribution}} = \frac{1}{Z_t} \underbrace{p(\mathbf{x}_t|\mathbf{Y}_{t-1})}_{\text{Predictive distribution}} \underbrace{l(\mathbf{y}_t|\mathbf{x}_t)}_{\text{Likelihood function}}$$

where Z_t is the normalizing constant defined by

$$Z_t = \int_{R^n} p(\mathbf{x}_t|\mathbf{Y}_{t-1}) l(\mathbf{y}_t|\mathbf{x}_t) d\mathbf{x}_t$$

Bayesian Filter (continued)

- **The celebrated Kalman filter is a special case of the Bayesian filter, assuming that the dynamic system is linear and both the dynamic noise and measurement noise are statistically independent processes.**

- **Except for this special case and couple of other cases, exact computation of the predictive distribution $p(\mathbf{x}_t | \mathbf{Y}_{t-1})$ is not feasible.**

- **We therefore have to abandon optimality and be content with a sub-optimal nonlinear filtering algorithm that is computationally tractable.**

Bayesian Filtering (continued)

Two Approaches for Approximate Nonlinear Filtering

1. Direct numerical approximation of the posterior in a local sense:
    - Extended Kalman filter (simple and therefore widely used)
    - Unscented Kalman filter
    - Central-difference Kalman filter
    - Cubature Kalman filter (New)

2. Indirect numerical approximation of the posterior in a global sense:
    - Particle filters:
      Roots embedded in Monte Carlo simulation
      Computationally demanding

Extended Kalman Filter

- **Linearize the system model around the filtered estimate $\hat{\mathbf{x}}_{t|t}$, and linearize the measurement model around the predicted estimate $\hat{\mathbf{x}}_{t|t-1}$**

- **Attributes and Limitations**

  (i) The EKF is simple to implement

  (ii) Estimation accuracy of the EKF is good for nonlinearities of a mild sort; otherwise, it is often *not* accurate enough.

9. The Cubature Kalman Filter

(Arasaratnam and Haykin, IEEE Trans. Automatic Control, accepted for publication, 2008)

- **At the heart of the Bayesian filter, we have to compute integrals whose integrands are expressed in the common form**

(Nonlinear function) x **(Gaussian function)**

- **The challenge is to numerically approximate the integral so as to completely preserve second-order information about the state $\mathbf{x}_t$ that is contained in the sequence of observations $\mathbf{y}_t$**

- **The computational tool that accommodates this requirement is the *cubature rule.***

Cubature Kalman Filter (continued)

The Cubature Rule

- **In mathematical terms, we have to compute an integral of the generic form**

$$h(\mathbf{f}) = \int_{R^n} \underbrace{\mathbf{f}(\mathbf{x})}_{\text{Arbitray nonlinear function}} \underbrace{\exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{x}\right)}_{\text{Normalized Gaussian function of zero mean and unit covariance matrix}} d\mathbf{x} \qquad (1)$$

- **To do the computation, the key step is to make a change of variables from the Cartesian coordinate system (in which the vector x is defined) to a spherical-radial coordinate system:**

$\mathbf{x} = r\mathbf{z}$ subject to $\mathbf{z}^T\mathbf{z} = 1$ and $\mathbf{x}^T\mathbf{x} = r^2$

where $0 \leq r < \infty$

Recursive Cycle of the Cubature Kalman Filter

- **The Kalman gain is computed as**

$$\mathbf{G}_t = \mathbf{P}_{xy,\, nt|nt\text{-}1}\mathbf{P}^{-1}_{yy,t|t\text{-}1}$$

where $\mathbf{P}^{-1}_{yy,t|t\text{-}1}$ is the inverse of the covariance matrix $\mathbf{P}_{yy,t|t\text{-}1}$.

- **Upon receiving the new observation $\mathbf{y}_t$, the filtered estimate of the state $\mathbf{x}_t$ is computed in accordance with the predictor-corrector formula:**

$$\underbrace{\hat{\mathbf{x}}_{t|t}}_{\text{Updated estimate}} = \underbrace{\hat{\mathbf{x}}_{t|t\text{-}1}}_{\text{Old estimate}} + \underbrace{\mathbf{G}_t}_{\text{Kalman gain}}\underbrace{(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t\text{-}1})}_{\text{Innovations process}}$$

- **Correspondingly, the covariance matrix of the filtered state estimation error is computed as**

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t\text{-}1} - \mathbf{G}_t\mathbf{P}_{yy,\, t|t\text{-}1}\mathbf{G}_t^T$$

Updated posterior distribution

$$p(\mathbf{x}_t|\mathbf{Y}_t) = R(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t})$$

Properties of the Cubature Kalman Filter

***Property* 1: The cubature Kalman filter (CKF) is a *derivative-free on-line sequential-state estimator*: It relies on integration for its operation.**

***Property* 2: Approximations of the moment integrals are all *linear* in the number of adjustable parameters.**

***Property* 3: Computational complexity of the cubature Kalman filter as a whole, grows as n^3, where n is the dimensionality of the state space.**

***Property* 4: The cubature Kalman filter *completely preserves second-order information* about the state that is contained in the observations.**

***Property* 5: The cubature Kalman filter *inherits properties of the linear Kalman filter*, including square-root filtering for improved accuracy and reliability.**

***Property* 6: The cubature Kalman filter is the closest known direct approximation to the Bayesian filter, outperforming the extended Kalman filter and the central-difference Kalman filter:**

It eases the curse-of-dimensionality problem
but, by itself, does not overcome it.

10. Supervised Learning

Training sample: $\{\mathbf{u}_t, \mathbf{d}_t\}$

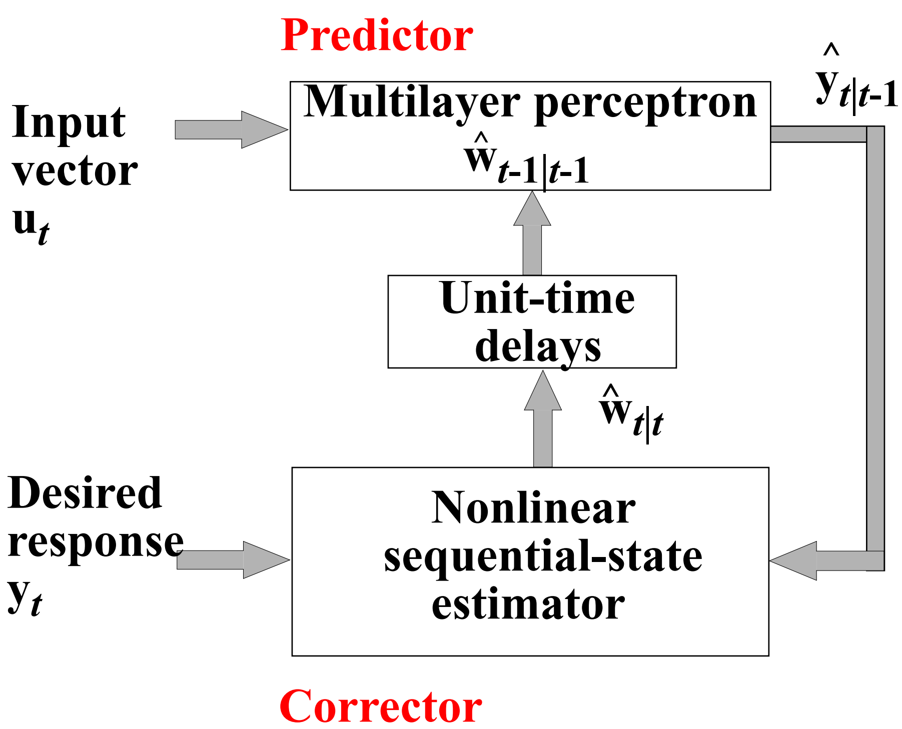

Figure 4: Block diagram of supervised learning machinery

Concluding Remark

"Cognitive Dynamic Systems"

are

A Way of the Future

in

The 21st Century

Concluding Remarks (continued)

Two New Books to watch out for:

1. **Neural Networks and Learning Machines**
   Simon Haykin
   Prentice-Hall, 3rd edition
   September 2008

2. **Foundations of Cognitive Dynamic Systems**
   Simon Haykin
   Cambridge University Press
   (In preparation)

Acknowledgements

- Many thanks to my gifted group of outstanding Ph.D. students

- Deep gratitude to the Natural Sciences and Engineering Research Council (NSERC) of Canada for continued financial support