Artificial Neural Networks

Connectionist Learning Machines

Mehmet K. Muezzinoglu

CECS492-01
Class Presentation

February 27, 2004
Outline

• Motivation
• History
• Algebraic models
• Recurrent models
• Applications
• Research trends
Biological Information Processing Units

Every organism is able to perceive and react to its habitat to an extent determined by its information processing capability.

The improvements on species’ neural systems has become dominant toward last steps of the evolution.

Information processing systems of vertebrates are made up of selective conductor cells, called *neurons*.
Mathematical Description of Neuro-Processing

Assumptions

- The net input to a neuron is a weighted sum of action potentials received from other neurons in physical contact.
- Each neuron performs a static mapping (activation function) on its net input to produce its own action potential.
  \[ y = \varphi(w^T \cdot x - t) \]
- Overall processing is organized as a cascading of a finite number of operational levels, which constitute the layers.
Artificial vs Natural Neuro-Processing

- Neurons fire in frequency domain. The response of its mathematical model is represented by magnitude.
- Natural signal processing can’t be categorized into layers. Thus canonizing the nervous system is a great challenge.
- Human nervous system utilizes $\sim 10^{11}$ reconfigurable units, where each neuron may take part in several tasks.

Originated by observations of natural signal processing, artificial neural networks literature today mostly concerns modelling and unifying the concept of learning, not necessarily by imitating the natural way.

A relatively small subset of studies are really classified under cognitive science, and report significant achievements.
Fundamental Discoveries

40’s to 60’s

- McCullough-Pitts Neuron (A logical calculus of the ideas immanent in nervous activity, 1943)

- Hebb’s postulate (The Organization of Behaviour, 1949)

When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A’s efficiency, as one of the cells firing B, is increased.
Fundamental Discoveries (cont’d)

• Rosenblatt’s perception model (The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, 1958)
Mark I Perceptron, the first machine that could “learn” to recognize and identify optical patterns.

• Adaptive Linear Element (ADALINE) and Widrow-Hoff Learning rule (1960)
Relatively Calm Years

70’s

- Questioning the capabilities of artificial neurons (Minsky and Papert, 1969)

- Self-organizing network (Kohonen, 1972)

- Adaptive Resonance Theory (Grossberg, 1976)

- The Brain-State-in-a-Box-Model (Anderson, 1978)
Development of Novel Tools

80’s

• The Hopfield Network (Hopfield, 1982)

• Error Back-Propagation Algorithm (Rumelhart & McClelland, 1986)

• NETTalk (Sejnowski, 1986)

• Hopfield-Tank Model (1987)

• Cellular Neural Networks (Chua & Wang, 1988)
Recognized as Superior Computing Tools

90’s

- Associative Memory

- Neuro-Control (Werbos, 1990) and Neuro-Dynamic Programming (Reinforcement Learning) (Bertsekas & Tsitsiklis, 1996)

- Regularization Framework for Neural Regressors (Poggio & Girosi, 1990)

- Support Vector Machines (Vapnik, 1995)
The Discrete Perceptron

This simple decision unit is able to implement “If-Then” rules.

First introduced by McCullough-Pitts to implement binary dichotomies.

\[ \sum \phi(\cdot) \]

\[ x_1, w_1, \ldots, x_n, w_n, t \in \{-1, 0, 1\}, \ t \in \{\pm0.5, \pm1.5\}, \ y \in \{0, 1\} \]

Due to the relaxation

\[ x_i, w_i, t \in \mathbb{R} \]

on the McCullough-Pitts model, any linearly separable dichotomy can be realized.
The Discrete Perceptron (cont’d)

Basic building blocks of Threshold Logic.

*Discrete Perceptron Training Algorithm* - A very simple and powerful learning procedure. Basically solves a linear inequality system.

Linearly separability is a very strict constraint, but worth checking when faced with a classification problem, e.g. determining edible gilled mushrooms of North America.
Multi-Layer Perceptron

Though there is no systematic design procedure yet, it has been proven that, a three-layer discrete perceptron can implement any dichotomy. (This can be verified easily for binary dichotomies.) Its utility is not limited to classification: By employing saturated, continuous activation functions, one can obtain efficient regressors. In particular, a three-layer network with sigmoidal activation functions possesses universal approximation capability.

\[ \text{Sampled Data } \xrightarrow{\text{sgm}(\text{sgm}(\text{sgm}(\cdot))))} \]
Error Back-Propagation Algorithm

Given the training set \( \{(x^i, y^i)\}_{i=1}^N \), the aim is to minimize

\[
E(W_1, W_2, W_3) = \sum_{i=1}^{N} \|y^i - \text{sgm}(W_3 \cdot \text{sgm}(W_2 \cdot \text{sgm}(W_1 \cdot x^i)))\|^2
\]

over \( W_1, W_2, W_3 \).

It is a gradient-descent optimization technique.
Radial Basis Function Networks

A hybrid topology consisting of linear and RBF neurons.

RBF neurons provide selective receptiveness.

RBFNs are universal approximators, too.

Determination of centers is the major challenge of training. Error back-propagation is still applicable, as well as clustering.
Self-Organizing Maps

Unsupervised learning has become popular in 70’s.

The goal is to render a network adapt itself to some undetermined features of the provided data.

If the data is presented sequentially to the network, then the organization is sensitive to the order of the incoming data.

It is also dependent upon the choice of initial condition.
Role of Feedback

Introducing feedback to an algebraic system brings many advantages:

- time-varying behavior
- simpler network architecture

as well as novel problems:

- stability
- control

Recurrent networks are able to handle efficiently time-varying signals, but their design is usually more complicated.
Brain-State-in-a-Box Model

A single-layer saturated-linear network with positive feedback operating in continuous time.

\[ x_1, x_n \rightarrow \sum w_{1n}, w_{nn}, w_{n1} \rightarrow \dot{x}_1, \dot{x}_n \]

Its unstable character forces the trajectory to infinity, while its activation functions restrict it within the unit hypercube.
Discrete Hopfield Model

- A single layer of discrete perceptrons operating in discrete time. (Hard-limiter type activation function).
- Its trajectory wanders on the finite set \( \{-1, 1\}^n \).
- Asynchronous or synchronous operation modes.
- Can be used to recall binary patterns.
- The recurrence
  \[
  x[k + 1] = \text{sgn} (W \cdot x[k] - t)
  \]
  minimizes a discrete quadratic
  \[
  Q(x) = -\frac{1}{2} \cdot x^T \cdot W \cdot x + t^T \cdot x.
  \]
Outer-Product Learning Rule

- A simple method to introduce fixed points to DHN.

\[ \mathbf{W} = \sum_{i=1}^{p} \mathbf{x}^i \cdot (\mathbf{x}^i)^T, \quad t = 0 \]

- No guarantee that each \( \mathbf{x}^i \) will become a fixed point.
- Attractiveness is not ensured.
- Many spurious (undesired) memories are introduced.
- The first method that inspired the neural associative memory.
- Better learning rules exist.
Associative Memory

Auto-association is a fundamental task performed by animal brain. Address addressable or content addressable memories. Not necessarily based on comparisons.

Mathematical formulation

\[
f (x^d) = \arg \min_{y \in M} d (x^d, y)
\]
Recurrence as a Problem Solver

Dynamical networks can perform deterministic search within unit hypercube.

The basic idea is to equate the network energy function to the cost of the problem at hand.

Hopfield and Tank proposed the first connectionist solution to the travelling salesman problem.

Good approximations for maximum clique problem have been reported recently.
Applications

One of the most widely-used computing tools, if not the first.

- *Pattern Recognition* (80% of applications): check reading; pap smear classification; character recognition; hand writing/envelope readers; explosives detection - JFK, HTR, Dallas airports; single-letter speech (spelling) recognition - British Telecom

- *Computer Vision*: satellite image compression - Royal Signal and Radar Est. UK

- *Diagnostics and Monitoring*: medical diagnostics - low back pain and coronary occlusion evaluation; helicopter gearbox failure detection; water analysis - Lyon, France; noisy motor diagnostics at Siemens
Applications (cont’d)

- *Process Control and Optimization*: automatic vehicle navigation; chemical batch process control via monitoring of pH levels; microchip fabrication process monitoring, control and centering; optimal dispatching of power at Tokyo Electric Power Co.

- *Expert Systems in Finances*: risk assessment evaluation; credit risk scoring, mortgage loan underwriting; prediction of currency exchange rates - Citibank; bond and stocks prices forecasting

- *Biomedical Signal Analysis*: cardiac signal compression - Siemens, discrimination of abnormalities in ECG, EEG waves
Attractive Features

• ANNs handle raw data and extract rules.

• More data typically implies better results.

• Efficiency in data representation.

• Easily implementable.

• Simulate learning.
Shortages

• Still far from explaining the natural intelligence.

• Sometimes too practical to maintain development.

• System theoretic approaches are required.

• Statistical explanations are necessary.
Notes on Research Trends

Statistical learning theory has dominated other theoretical approaches in the last 5 years.

Control applications has been attracting increasing attention recently.

Interdisciplinary studies, especially with cognitive scientists, are promising.
### LIST OF ISI JOURNALS

**CATEGORY : COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE**

**(FIRST 20 of 72)**

Sorted by: Impact Factor

<table>
<thead>
<tr>
<th>Abbreviated Journal Title</th>
<th>ISSN</th>
<th>2002 Total Cites</th>
<th>Impact Factor</th>
<th>Immediacy Index</th>
<th>2002 Articles</th>
<th>Cited Half-life</th>
</tr>
</thead>
<tbody>
<tr>
<td>J MACH LEARN RES</td>
<td>1532-4435</td>
<td>50</td>
<td>3.818</td>
<td>0.267</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>IEEE T PATTERN ANAL</td>
<td>0162-8828</td>
<td>8835</td>
<td>2.923</td>
<td>0.212</td>
<td>146</td>
<td>9.2</td>
</tr>
<tr>
<td>MED IMAGE ANAL</td>
<td>1361-8415</td>
<td>547</td>
<td>2.680</td>
<td>0.179</td>
<td>28</td>
<td>4.5</td>
</tr>
<tr>
<td>COGNITIVE BRAIN RES</td>
<td>0926-6410</td>
<td>1261</td>
<td>2.404</td>
<td>0.275</td>
<td>91</td>
<td>4.1</td>
</tr>
<tr>
<td>NEURAL COMPUT</td>
<td>0899-7667</td>
<td>3443</td>
<td>2.313</td>
<td>0.325</td>
<td>120</td>
<td>6.2</td>
</tr>
<tr>
<td>INT J COMPUT VISION</td>
<td>0920-5691</td>
<td>2321</td>
<td>2.034</td>
<td>0.250</td>
<td>68</td>
<td>7.4</td>
</tr>
<tr>
<td>MACH LEARN</td>
<td>0885-6125</td>
<td>2087</td>
<td>1.944</td>
<td>0.852</td>
<td>54</td>
<td>7.3</td>
</tr>
<tr>
<td>IEEE INTELL SYST</td>
<td>1094-7167</td>
<td>449</td>
<td>1.905</td>
<td>0.286</td>
<td>42</td>
<td>3.1</td>
</tr>
<tr>
<td>CHEMOMETR INTELL LAB</td>
<td>0169-7439</td>
<td>1703</td>
<td>1.841</td>
<td>0.182</td>
<td>77</td>
<td>5.7</td>
</tr>
<tr>
<td>ARTIF INTELL</td>
<td>0004-3702</td>
<td>4040</td>
<td>1.769</td>
<td>0.453</td>
<td>75</td>
<td>&gt;10.0</td>
</tr>
<tr>
<td>NETWORK-COMP NEURAL</td>
<td>0954-898X</td>
<td>478</td>
<td>1.733</td>
<td>0.577</td>
<td>26</td>
<td>6.5</td>
</tr>
<tr>
<td>NEURAL NETWORKS</td>
<td>0893-6080</td>
<td>2626</td>
<td>1.557</td>
<td>0.341</td>
<td>88</td>
<td>8.3</td>
</tr>
<tr>
<td>ARTIF LIFE</td>
<td>1064-5462</td>
<td>401</td>
<td>1.526</td>
<td>0.385</td>
<td>13</td>
<td>7.5</td>
</tr>
<tr>
<td>IEEE T EVOLUT COMPUT</td>
<td>1089-778X</td>
<td>400</td>
<td>1.486</td>
<td>0.024</td>
<td>41</td>
<td>3.9</td>
</tr>
<tr>
<td>IEEE T NEURAL NETWOR</td>
<td>1045-9227</td>
<td>3565</td>
<td>1.404</td>
<td>0.112</td>
<td>143</td>
<td>7.1</td>
</tr>
<tr>
<td>AI MAG</td>
<td>0738-4602</td>
<td>536</td>
<td>1.328</td>
<td>0.031</td>
<td>32</td>
<td>6.3</td>
</tr>
<tr>
<td>IEEE T FUZZY SYST</td>
<td>1063-6706</td>
<td>1443</td>
<td>1.324</td>
<td>0.043</td>
<td>69</td>
<td>6.0</td>
</tr>
<tr>
<td>DATA MIN KNOWL DISC</td>
<td>1384-5810</td>
<td>300</td>
<td>1.192</td>
<td>0.786</td>
<td>14</td>
<td>4.6</td>
</tr>
<tr>
<td>AUTON AGENT MULTI-AG</td>
<td>1387-2532</td>
<td>92</td>
<td>1.129</td>
<td>0.250</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>IEEE T KNOWL DATA EN</td>
<td>1041-4347</td>
<td>1073</td>
<td>1.055</td>
<td>0.113</td>
<td>97</td>
<td>5.9</td>
</tr>
</tbody>
</table>