Computing Like the Brain
The Path To Machine Intelligence

Jeff Hawkins
GROK - Numenta
jhawkins@groksolutions.com
1) Discover operating principles of neocortex

2) Build systems based on these principles
Artificial Intelligence - no neuroscience

Alan Turing

“Computers are universal machines” 1935
“Human behavior as test for machine intelligence” 1950

Major AI Initiatives
- MIT AI Lab
- 5th Generation Computing Project
- DARPA Strategic Computing Initiative
- DARPA Grand Challenge

AI Projects
- ACT-R
- Asimo
- CoJACK
- Cyc
- Deep Blue
- Global Workspace Theory
- Mycin
- SHRDLU
- Soar
- Watson
- Many more -

Pros: - Good solutions
Cons: - Task specific
- Limited or no learning
Artificial Neural Networks – minimal neuroscience

Warren McCulloch
Walter Pitts

“Neurons as logic gates” 1943
Proposed first artificial neural network

ANN techniques
- Back propagation
- Boltzman machines
- Hopfield networks
- Kohonen networks
- Parallel Distributed Processing
- Machine learning
- Deep Learning

Pros:
- Good classifiers
- Learning systems

Cons:
- Limited
- Not brain like
Whole Brain Simulator – maximal neuroscience

The Human Brain Project

No theory

No attempt at Machine Intelligence
1) Discover operating principles of neocortex

2) Build systems based on these principles

Good progress is being made

1940s in computing = 2010s in machine intelligence
The neocortex is a memory system.

The neocortex learns a model from sensor data
- predictions
- anomalies
- actions

The neocortex learns a sensory-motor model of the world
1) On-line learning from streaming data
1) On-line learning from streaming data

2) Hierarchy of memory regions
   - regions are nearly identical
Principles of Neocortical Function

1) On-line learning from streaming data
2) Hierarchy of memory regions
3) Sequence memory
   - inference
   - motor
Principles of Neocortical Function

1) On-line learning from streaming data
2) Hierarchy of memory regions
3) Sequence memory
4) Sparse Distributed Representations
Principles of Neocortical Function

1) On-line learning from streaming data
2) Hierarchy of memory regions
3) Sequence memory
4) Sparse Distributed Representations
5) All regions are sensory and motor
Principles of Neocortical Function

1) On-line learning from streaming data
2) Hierarchy of memory regions
3) Sequence memory
4) Sparse Distributed Representations
5) All regions are sensory and motor
6) Attention
Principles of Neocortical Function

1) On-line learning from streaming data
2) Hierarchy of memory regions
3) Sequence memory
4) Sparse Distributed Representations
5) All regions are sensory and motor
6) Attention

These six principles are necessary and sufficient for biological and machine intelligence.

- All mammals from mouse to human have them
- We can build machines like this
**Sparse Distributed Representations (SDRs)**

- Many bits (thousands)
- Few 1’s mostly 0’s
- Example: 2,000 bits, 2% active
  
  \[
  0100000000000000000000000000000000000000000000000000000000000000000
  \]
- Each bit has semantic meaning (learned)
- Representation is semantic

**Dense Representations**

- Few bits (8 to 128)
- All combinations of 1’s and 0’s
- Example: 8 bit ASCII
  
  \[
  01101101 \text{ = } m
  \]
- Individual bits have no inherent meaning
- Representation is arbitrary
**SDR Properties**

1) **Similarity:**
   - shared bits = semantic similarity

2) **Store and Compare:**
   - store indices of active bits
   - subsampling is OK

3) **Union membership:**
   - Is this SDR a member?
Sequence Memory (for inference and motor)

How does a layer of neurons learn sequences?

Coincidence detectors
Each cell is one bit in our Sparse Distributed Representation (SDR).

SDRs are formed via a local competition between cells.

All processes are local across large sheets of cells.
SDR (time = 1)
SDR (time = 2)
Cells connect to sample of previously active cells to predict their own future activity.
Multiple Predictions Can Occur at Once.

This is a 1\textsuperscript{st} order memory.
We need a high order memory.
High order sequences are enabled with multiple cells per column.
High Order Sequence Memory

40 active columns, 10 cells per column

= $10^{40}$ ways to represent the same input in different contexts

A-B-C-D-E
X-B’-C’-D’-Y
High Order Sequence Memory

Distributed sequence memory
High order, high capacity
Noise and fault tolerant
Multiple simultaneous predictions
Semantic generalization
Online learning

- Learn continuously, no batch processing
- If pattern repeats, reinforce, otherwise forget it

Learning is the growth of new synapses.

Connection strength is binary
Connection permanence is a scalar
Training changes permanence
"Cortical Learning Algorithm" (CLA)

Not your typical computer memory!
A building block for
- neocortex
- machine intelligence
Evidence suggests each layer is implementing a CLA variant
What Is Next? Three Current Directions

1) Commercialization
   - GROK: Predictive analytics using CLA
   - Commercial value accelerates interest and investment

2) Open Source Project
   - NuPIC: CLA open source software and community
   - Improve algorithms, develop applications

3) Custom CLA Hardware
   - Needed for scaling research and commercial applications
   - IBM, Seagate, Sandia Labs, DARPA
GROK: Predictive Analytics Using CLA

**Encoders**
Convert native data type to SDRs

**CLA**
Learns spatial/temporal patterns
Outputs
- predictions
anomalies

**Sequence Memory**
2,000 cortical columns
60,000 neurons
- variable order
- online learning

**SDRs**

**Predictions Anomalies**

**Actions**

Field 1   Field 2   Field 3   Field N
Field 1   Field 2   Field 3   Field N
Field 1   Field 2   Field 3   Field N

numbers
categories
text
date
time
GROK example: Factory Energy Usage

![Energy Usage Graph](image_url)

- Energy Usage (Kw)
- Dates: 26-Feb to 6-Mar
Customer need

At midnight, make 24 hourly predictions
GROK Predictions and Actuals
GROK example: Predicting Server Demand

Grok used to predict server demand

Approximately 15% reduction in AWS cost

Server demand, Actual vs. Predicted
GROK example: Detecting Anomalous Behavior

Grok builds model of data, detects changes in predictability.

Gear bearing temperature & Grok Anomaly Score

GROK going to market for anomaly detection in I.T. 2014
2) Open Source Project

NuPIC: www.Numenta.org
- CLA source code (single tree), GPLv3
- Papers, videos, docs

Community
- 200+ mail list subscribers, growing
- 20+ messages per day
- full time manager, Matt Taylor

What you can do
- Get educated
- New applications for CLA
- Extend CLA: robotics, language, vision
- Tools, documentation

2nd Hackathon November 2,3 in San Francisco
- Natural language processing using SDRs
- Sensory-motor integration discussion
- 2014 hackathon Ireland?
3) Custom CLA Hardware

**HW companies looking “Beyond von Neumann”**
- Distributed memory
- Fault tolerant
- Hierarchical

**New HW Architectures Needed**
- Speed (research)
- Cost, power, embedded (commercial)

**IBM**
- Almaden Research Labs
- Joint research agreement

**DARPA**
- New Program called “Cortical Processor”
- HTM (Hierarchical Temporal Memory)
- CLA is prototype primitive

**Seagate**

**Sandia Labs**
Future of Machine Intelligence
Future of Machine Intelligence

**Definite**
- Faster, Bigger
- Super senses
- Fluid robotics
- Distributed hierarchy

**Maybe**
- Humanoid robots
- Computer/Brain interfaces for all

**Not**
- Uploaded brains
- Evil robots
Why Create Intelligent Machines?

Live better

Learn more

Thank You