Carnegie Mellon University Africa
Certificate Program in AI and Machine Learning in Africa

Course AIML01: Artificial Intelligence – Past, Present, and Future

Module 2: The Nature of AI
Lecture 2: Connectionist AI – From Perceptrons to Deep Neural Networks

Welcome to Lecture 2 of Module 2. We continue to look more deeply at the technical foundations of AI and machine learning, focusing here on connectionist AI.

In this lecture, we explain how connectionist AI has its roots in early work psychology and associative memory. We explore the early implementation of connectionist AI using elementary artificial neural networks, and we trace their development over the past eighty years, culminating in modern high performance deep machine learning.

There are five learning objectives, so that, after studying the material covered in this lecture, you should be able to do the following.

1. Explain how connectionist systems and artificial neural networks represent and process information.

2. Explain the operation, strengths, and weaknesses of perceptron architectures.

3. Explain how modern artificial networks achieve their high performance.

4. Identify different types of artificial neural network and the types of application for which they are best suited.

Slide 1 Welcome to Lecture 2 of Module 2. We continue to look more deeply at the technical foundations of AI and machine learning, focusing here on connectionist AI.

Slide 2 Connectionist AI differs from symbolic AI in that information is represented in non-symbolic form (e.g., by an image or a sound) and it is processed by propagating it through an interconnected network of relatively simple processing elements, typically implemented as artificial neural networks.

They use statistical properties rather than logical rules to process data and extract the required information

In the following, we will summarize the main principles of connectionism, briefly tracing its history and highlighting the main developments that have led us to where we are today.

Unfortunately, but inevitably, we will be forced into making use of many technical terms with little or no explanation

Still, we will give some sense of the essence of connectionism, and we will follow up in other courses to deal with these concepts in greater depth.

Slide 3 The term *connectionist model* is usually attributed to Feldman and Ballard in 1982, Feldman used the term the previous year in a book on associative memory.

Slide 4 However, the roots of connectionism reach back well before the computational era,

with connectionist principles clearly evident in William James' nineteenth century model of associative memory (James, 1890).

Slide 5 An associative memory allows patterns of information to be stored and recalled based on the presentation of other patterns.

Slide 6 The pattern that is recalled can be similar to the pattern that is used to recall it

For example, a part of a visual image would recall the complete image

Slide 7 Or a degraded version of it would recall the original image.

Slide 8 The input and recalled patterns can be different types or modalities. For example, a scent or sound might recall an associated image.

Slide 9 James's work also anticipated Hebbian learning,

an influential unsupervised neural training process whereby the synaptic strength — the bond between connecting neurons — is increased if both the source and target neurons are active at the same time.

This is sometimes capture in the phrase "neurons that fire together, wire together".

Prior to James, in 1873 Alexander Bain also contributed the idea of Hebbian Learning, linking the process of associative memory with the distribution of activity of neural groupings, his term for neural networks.

Slide 10 The introduction to Donald Hebb's book *The Organization of Behaviour* published in 1949 also contains one of the first usages of the term connectionism.

Slide 11 The seminal paper by Warren S. McCulloch and Walter Pitts in 1943

"A logical calculus immanent in nervous activity"

is also regarded as the foundation of artificial neural networks and connectionism.

Slide 12 They showed that any statement within propositional logic could be represented by a network of simple processing units, i.e., a connectionist system.

They also showed that such nets have, in principle, the computational power of a Universal Turing Machine, the theoretical basis for all computation.

Slide 13 The connectionist approach was advanced significantly in the late 1950s with the introduction of Frank Rosenblatt's perceptron.

Slide 14 Which processes a sum of weighted input feature values

Slide 15 One perceptron can perform binary classification; that is, it can distinguish between two classes (provided their features can be separated by a straight line).

Slide 16 Several perceptrons can distinguish between several classes

Slide 17 Again, provided their features can be separated by straight lines.

Slide 18 The perceptron was originally intended as a machine, not as an algorithm or program.

The weights encoded by arrays of potentiometers which were turned by electric motors during training

Slide 19 Although network learning advanced in 1960 with the introduction of the Widrow-Hoff rule (also called the delta rule) for supervised training

Perceptron networks suffered from a severe problem: no learning algorithm existed to allow the adjustment of the weights of the connections between input units and hidden associative units in networks with more than two layers, i.e., multi-layered perceptrons (MLPs).

Slide 20 In 1969, Marvin Minsky and Seymour Papert caused something of a stir by showing that perceptron networks can only be trained to solve linearly separable problems.

And couldn't be trained to solve more general problems.

This had a very negative influence on neural network research for over a decade.

Slide 21 As a result, research on neural networks and connectionism suffered considerably and marked the beginning of a decade-long winter for connectionist AI.

Slide 22 During the period that followed this disenchantment with perceptron networks, alternative connectionist models were developed, such as

adaptive resonance theory (ART), first introduced by Stephen Grossberg in 1976 and developed in the ensuing years (Carpenter and Grossberg, 1995), and

Teuvo Kohonen's self-organizing maps (Kohonen, 1982).

ART addresses real-time supervised and unsupervised category learning, pattern classification, and prediction,

while Kohonen networks exploit self-organization for unsupervised learning, and they can be used as either an auto-associative memory or a pattern classifier.

Slide 23 Perceptron-based neural networks underwent a strong resurgence in the mid-1980s with the introduction of the back-propagation algorithm by David Rumelhart, Geoff Hinton, and Ronald Williams

which had previously been derived independently by others

Backpropagation finally made it feasible to train MLPs, overcoming the restriction highlighted by Minsky and Papert (1969),

thereby enabling MLPs to learn solutions to complex problems that are not linearly separable.

This was a major breakthrough in neural network and connectionist research.

Slide 24 Perceptron-based neural networks typically represent a static mapping between the inputs and outputs in which data flows in just one direction through the network, from input to output.

There is an alternative, however, in which the network has connections that loop back to form cycles, that is, networks in which either the output or the hidden unit activation signals are fed back to the network as inputs.

These are called recurrent neural networks.

The recurrent pathways in the network introduce a dynamic, time-varying, behavior into the network operation

Slide 25 In 1982, John Hopfield introduced a fully-connected recurrent neural network model of associative memory: the Hopfield net

This was inspired by a model of human memory devised by William Little in 1974, based on the Ising model (also known as the Ising-Lenz model), a mathematical model of ferromagnetism in statistical mechanics.

Slide 26 Boltzmann machines are variants of Hopfield nets that use stochastic rather than deterministic weight update procedures to avoid problems with the network becoming trapped in local minima during learning.

Slide 27 However, in the late 1990s, significant breakthroughs in deep networks heralded a new era in connectionism.

These breakthroughs included long short-term memory (LSTM) by Sepp Hochreiter and Jürgen Schmidhuber in 1997 …

a form of recurrent neural network that can select which information to remember or forget when processing a sequence of data or time-varying information.

Slide 28 … and convolutional neural networks (CNNs) by Yann LeCun et al. (1998)

A CNN network is similar in principle to the multi-layer perceptrons of the 1980s and early 1990s, but they have more layers, each of which performs a different function.

In a CNN, convolution refers to the application of a filter to the data being processed by the neural network.

The key feature of a CNN is that these filters are learned by the network during the training phase.

This marked a significant departure from previous approaches where the filters, and the features they extracted, were the result of hand-crafted design.

Consequently, CNNs are able to map directly from the input space, e.g., the image to be classified or the image in which you want to search for a given object, directly to the image label or the object location.

For this reason, they are referred to as end-to-end systems.

Slide 29 However, it took another ten years before they were widely adopted because of the lack of sufficiently large data sets and sufficient computational power for training.

In 2011, AlexNet (Krizhevsky et al., 2012), a CNN with seven hidden layers won the ImageNet Large Scale Visual Recognition Challenge, using 1.2 million training images for 1000 classes.

Slide 30 What happened in the meantime?

There was a second AI winter

followed by a neural network spring with the arrival of

(a) Massive, labelled data sets for training

(b) Huge increase in computer power in the form of graphic processor units (GPU) to facilitate training in reasonable amount of time

Prior to this, hardware was slow for floating point computation and training a character recognizer took 2 weeks on a Sun or SGI workstation

Slide 31 For the past ten years, we have been enjoying a deep learning summer ...

Slide 32 Deep convolutional neural networks, such as VGG, GoogLeNet (now called more commonly referred to as Inception), ResNet, and DenseNet, have been applied successfully in many challenging applications.

The networks have 17, 19, 22, or many more layers.

Slide 33 and performance has improved through the use new network elements and new techniques for learning, such as

- more effective activation functions (e.g. the rectified linear unit ReLU),
- the use of specialized layers (e.g. pooling),
- more advanced learning techniques (e.g. batch normalization and dropout),
- techniques to overcome the problem of vanishing gradients (where the error terms become too small to produce an improvement in network performance as they are propagated back in a deep network)
- and a better understanding of how to adjust the system hyper-parameters during training to improve performance.

Slide 34 While the Imagenet challenge, which was concerned only with classification of images, provided the impetus for many of these advances,

computer vision, which is concerned with the more general goal inferring the content, organization, and behaviour of a 3D world by the automatic analysis of images of that world, continued to drive the development of deep learning.

For example, CNNs and regional CNNs (RCNNs) have demonstrated impressive performance in object detection, recognition, and localization, face detection, face recognition, and object tracking.

Slide 35 Even in cluttered scenes.

Slide 36 New forms of recurrent neural networks proved very successful solving problems that involve processing and analysing sequences of states, e.g., in natural language,

by exploiting new recurrent elements such as long short term memory (LSTM) and gated recurrent units (GRU).

Slide 37 Progress has continued, with modern architectures successfully combining the power of deep CNNs and LSTMs to address problems that involve both images and language, e.g., automatic image annotation and captioning, image retrieval and synthesis base on linguistic description

Slide 38 The advent of generative adversarial networks, or GANs, which work as actor-critic systems, has provided the means for two learning networks to learn from each other and thereby improve the performance of both (Goodfellow et al., 2014; Goodfellow, 2017)

Slide 39 This has yielded remarkable results in image synthesis, among many other applications.

Slide 40 In 2019, Yann LeCun, introduced the idea of self-supervised learning, augmenting the traditional unsupervised, supervised, and reinforcement learning paradigms.

The key insight in self-supervised learning is that the system learns to predict parts of the data set that have been intentionally removed.

This means that it doesn't need a supervisor: the data provides the supervision.

Slide 41 Advances in deep learning continue to be made, with new techniques and architectures, such as

Autoencoder architectures (which use encoders to learn very efficient representations of data and decoders to interpret data)

Attention-based techniques that allow the neural network to select the most important elements in its processing pipeline

Transformer architectures which exploit attention

And hybrid architectures, sometimes referred to as neuro-symbolic architectures, that provide a blend of symbolic and subsymbolic processing,

among many others.

Slide 42 The Transformer, introduced by Google in 2017, has been particularly influential in recent years.

It replaces recurrence with attention mechanisms, allowing for significantly more parallelization than other models such as recurrent neural networks.

Slide 43 Tranformer architectures provided the basis for significant progress with deep neural networks for language understanding and generation.

Examples include BERT (for Bidirectional Encoder Representations from Transformers), LaMDA, and Gemini from Google,

the series of Generative Pre-trained Transformer (GPT) architectures from OpenAI, culminating, for now, in GPT-4,

and Llama 3 from Meta.

Trained on approximately 10 trillion of words at a cost of $100 million, and with over a trillion machine learning parameters, GPT-4 is capable of generating natural language text that is often indistinguishable from that generated by humans.

LaMDA's conversational abilities are so convincing that they have fooled some people into thinking that it is sentient.

Slide 44 Neural network models such as BERT, GPT-4, and LaMDA 2 are instances of a new approach to deep learning referred to as Foundation Models (Bommasani et al., 2021).

Natural language processing foundation models such BERT, GPT, and LaMDA, are frequently referred to as large language models (LLMs).

Slide 45 DALL-E 3, developed by OpenAI, is transformer-based text-to-image system based on GPT-3.

It can generate highly realistic digital images based on natural language descriptions of the desired content.

This image was generated from the text "An astronaut riding a horse in a photorealistic style".

DALL-E is derived from the name of artist Salvador Dali and the animated robot WALL-E.

Other AI-based image generator include Stable Diffusion from Stability AI, and Midjourney from Midjourney Inc.

Slide 46 Foundation models are also being used to program robots.

The Robotic Transformers, also referred to as Vision-Language-Action models, include

RT-1, RT-2, and RT-X from Google DeepMind,

and PaLM-E from Robotics at Google.

Today, this technology, based on foundation models, is generically referred to as "Generative AI".

To summarize:

- Connectionist AI has its roots in early work psychology and associative memory.

- Connectionist AI is typically implemented using artificial neural networks.

- Although limited at first, artificial neural networks for the basis of modern high performance deep machine learning.

- These modern artificial networks achieve their performance by using many layers of processing, very large training data sets, and very powerful GPU-based computers during the training phase.

Recommended Reading

Here is some recommended reading.

P. Bassereau, A. Cangelosi, J. Cejkova, C. Gershenson, R. Goldstein, Z. Martins, S. Matthews, F. Ritort, S. Seager, B. Van Tiggelen, D. Vernon, and F. Westall, Chapter 4 Physics for Understanding Life, **Section 4.3 Artificial Intelligence: Powering the Fourth Industrial Revolution**, EPS Grand Challenges: Physics for Society at the Horizon 2050, IOP Publishing, 2024.
http://vernon.eu/publications/2024_Bassereau_et_al.pdf

van Veen, F. The Neural Network Zoo, The Asimov Institute.
https://www.asimovinstitute.org/neural-network-zoo/

Wang, H. and Raj, B. (2017). On the Origin of Deep Learning, arXiv:1702.07800v4.
https://arxiv.org/pdf/1702.07800.pdf

References

Here are the references cited to support the main points in what we covered today.

Bommasani, R. et al. (66 authors), On the Opportunities and Risks of Foundation Models}, arxiv 2108.07258, 2021.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G, Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D. (2020) Language models are few-shot learners. arXiv:2005.14165.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Hebb, D. O. *The Organization of Behaviour – A neuropsychological theor*y. John Wiley & Sons, New York, 1949. https://pure.mpg.de/rest/items/item_2346268_3/component/file_2346267/content

Hopfield, J. J. Neural networks and physical systems with emergent collective computational properties. Proc. Nat. Acad. Sci. (USA), 79, 2554–2558, 1982.

James, W. *The Principles of Psychology*, Volume 1. Harvard University Press, Cambridge, MA, 1890.

LeCun, Y. Deep Learning Hardware: Past, Present, & Future, ISSCC 2019

McCulloch, W. S. and Pitts, W. A logical calculus of ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5:115–133, 1943.

Feldman, J. A. and Ballard, D. H. "Connectionist models and their properties," Cognitive Science, 6,205-254, 1982.

Feldman, J. A. "A connectionist model of visual memory," in Parallel Models of Associative Memory, G.E. Hinton and J.A. Anderson (eds.), Lawrence Erlbaum Associates, Inc., Publishers, Hillsdale NJ, 1981.

Little, W. The existence of persistent states in the brain. Mathematical Biosciences, 19:101-120, 1974.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, NIPS 2017, Vol. 20.