SNS COLLEGE OF TECHNOLOGY
(An Autonomous Institution)
Coimbatore – 35.

DEPARTMENT OF BIOMEDICAL ENGINEERING

UNIT 2

History of Deep Learning

Deep Learning, is a more evolved branch of machine learning, and uses layers of algorithms to process data, and imitate the thinking process, or to develop *abstractions*.

It is often used to visually recognize objects and understand human speech. Information is passed through each layer, with the output of the previous layer providing input for the next layer. The first layer in a network is called the input layer, while the last is called an output layer.

All the layers between input and output are referred to as hidden layers. Each layer is typically a simple, uniform algorithm containing one kind of activation function.

Feature extraction is another aspect of deep learning. It is used for pattern recognition and image processing. Feature extraction uses an algorithm to automatically construct meaningful "features" of the data for purposes of training, learning, and understanding. Normally a data scientist, or a programmer, is responsible for feature extraction.

The history of deep learning can be traced back to 1943, when Walter Pitts and Warren McCulloch created a computer model based on the neural networks of the human brain.

They used a combination of algorithms and mathematics they called "threshold logic" to mimic the thought process. Since that time, Deep Learning has evolved steadily, with only two significant breaks in its development. Both were tied to the infamous Artificial Intelligence winters.

The 1960s

Henry J. Kelley is given credit for developing the basics of a continuous Back Propagation Modelin 1960. In 1962, a simpler version based only on the chain rule was developed by

Stuart Dreyfus. While the concept of back propagation (the backward propagation of errors for purposes of training) did exist in the early 1960s, it was clumsy and inefficient, and would not become useful until 1985.

The earliest efforts in developing deep learning algorithms came from Alexey Grigoryevich Ivakhnenko (developed the *Group Method of Data Handling*) and Valentin Grigorʹevich Lapa (author of *Cybernetics and Forecasting Techniques)* in 1965. They used models with polynomial (complicated equations) activation functions, that were then analyzed statistically. From each layer, the best statistically chosen features were then forwarded on to the next layer (a slow, manual process).

The 1970s

During the 1970's the first AI winter kicked in, the result of promises that couldn't be kept. The impact of this lack of funding limited both DL and AI research. Fortunately, there were individuals who carried on the research without funding.

The first "convolutional neural networks" were used by Kunihiko Fukushima. Fukushima designed neural networks with multiple pooling and convolutional layers. In 1979, he developed an artificial neural network, called Neocognitron, which used a hierarchical, multilayered design. This design allowed the computer the "learn" to recognize visual patterns. The networks resembled modern versions, but were trained with a reinforcement strategy of recurring activation in multiple layers, which gained strength over time. Additionally, Fukushima's design allowed important features to be adjusted manually by increasing the "weight" of certain connections.

Many of the concepts of Neocognitron continue to be used.

The use of top-down connections and new learning methods have allowed for a variety of neural networks to be realized. When more than one pattern is presented at the same time, the Selective Attention Model can separate and recognize individual patterns by shifting its attention from one to the other. (The same process many of us use when multitasking). A modern Neocognitron can not only identify patterns with missing information (for example, an incomplete number 5), but can also complete the image by adding the missing information. This could be described as "inference."

Back propagation, the use of errors in training deep learning models, evolved significantly in 1970. This was when Seppo Linnainmaa wrote his master's thesis, including a FORTRAN code for back propagation.

Unfortunately, the concept was not applied to neural networks until 1985. This was when Rumelhart, Williams, and Hinton demonstrated back propagation in a neural network could provide "interesting" distribution representations. Philosophically, this discovery brought to light the question within cognitive psychology of whether human understanding relies on symbolic logic (computationalism) or distributed representations (connectionism).

The 1980s and 90s

In 1989, Yann LeCun provided the first practical demonstration of backpropagation at Bell Labs. He combined convolutional neural networks with back propagation onto read "handwritten" digits. This system was eventually used to read the numbers of handwritten checks.

This time is also when the second AI winter (1985-90s) kicked in, which also effected research for neural networks and deep learning. Various overly-optimistic individuals had exaggerated the "immediate" potential of Artificial Intelligence, breaking expectations and angering investors. The anger was so intense, the phrase Artificial Intelligence reached pseudoscience status. Fortunately, some people continued to work on AI and DL, and some significant advances were made. In 1995, Dana Cortes and Vladimir Vapnik developed the support vector machine (a system for mapping and recognizing similar data). LSTM (long short-term memory) for recurrent neural networks was developed in 1997, by Sepp Hochreiter and Juergen Schmidhuber.

The next significant evolutionary step for deep learning took place in 1999, when computers started becoming faster at processing data and GPU (graphics processing units) were developed. Faster processing, with GPUs processing pictures, increased computational speeds by 1000 times over a 10 year span. During this time, neural networks began to compete with support vector machines. While a neural network could be slow compared to a support vector machine, neural networks offered better results using the same data. Neural networks also have the advantage of continuing to improve as more training data is added.

2000-2010

Around the year 2000, *The Vanishing Gradient Problem* appeared. It was discovered "features" (lessons) formed in lower layers were not being learned by the upper layers, because no learning signal reached these layers. This was not a fundamental problem for all neural networks, just the ones with gradient-based learning methods. The source of the problem turned out to be certain activation functions. A number of activation functions condensed their input, in turn reducing the output range in a somewhat chaotic fashion. This produced large areas of input mapped over an extremely small range. In these areas of input, a large change will be reduced to a small change in the output, resulting in a vanishing gradient. Two solutions used to solve this problem were layer-by-layer pre-training and the development of long short-term memory.

In 2001, a research report by META Group (now called Gartner) described he challenges and opportunities of data growth as three-dimensional. The report described the increasing volume of data and the increasing speed of data as increasing the range of data sources and types. This was a call to prepare for the onslaught of Big Data, which was just starting.

In 2009, Fei-Fei Li, an AI professor at Stanford launched ImageNet, assembled a free database of more than 14 million labeled images. The Internet is, and was, full of unlabeled images. Labeled images were needed to "train" neural nets. Professor Li said, "Our vision was that big data would change the way machine learning works. Data drives learning."

2011-2020

By 2011, the speed of GPUs had increased significantly, making it possible to train convolutional neural networks "without" the layer-by-layer pre-training. With the increased computing speed, it became obvious deep learning had significant advantages in terms of efficiency and speed. One example is AlexNet, a convolutional neural network whose architecture won several international competitions during 2011 and 2012. Rectified linear units were used to enhance the speed and dropout.

Also in 2012, Google Brain released the results of an unusual project known as *The Cat Experiment*. The free-spirited project explored the difficulties of "unsupervised learning." Deep learning uses "supervised learning," meaning the convolutional neural net is trained using labeled data (think images from ImageNet). Using unsupervised learning, a convolutional neural net is given unlabeled data, and is then asked to seek out recurring patterns.

The Cat Experiment used a neural net spread over 1,000 computers. Ten million "unlabeled" images were taken randomly from YouTube, shown to the system, and then the training software was allowed to run. At the end of the training, one neuron in the highest layer was found to respond strongly to the images of cats. Andrew Ng, the project's founder said, "We also found a neuron that responded very strongly to human faces." Unsupervised learning remains a significant goal in the field of deep learning.

The Generative Adversarial Neural Network (GAN) was introduced in 2014. GAN was created by Ian Goodfellow. With GAN, two neural networks play against each other in a game. The goal of the game is for one network to imitate a photo, and trick its opponent into believing it is real. The opponent is, of course, looking for flaws. The game is played until the near perfect photo tricks the opponent. GAN provides a way to perfect a product (and has also begun being used by scammers).

The Future of Deep learning and Business

Deep learning has provided image-based product searches – Ebay, Etsy– and efficient ways to inspect products on the assembly line. The first supports consumer convenience, while the second is an example of business productivity.

Currently, the evolution of artificial intelligenceis dependent on deep learning. Deep learning is still evolving and in need of creative ideas.

Semantics technology is being used with deep learning to take artificial intelligence to the next level, providing more natural sounding, human-like conversations.

Banks and financial services are using deep learning to automate trading, reduce risk, detect fraud, and provide AI/chatbot advice to investors. A report from the EIU (Economist Intelligence Unit) suggests 86% of financial services are planning to increase their artificial intelligence investments by 2025.

Deep learning and artificial intelligence are influencing the creation of new business models. These businesses are creating new corporate cultures that embrace deep learning, artificial intelligence, and modern technology.

Reference:

https://www.dataversity.net/brief-history-deep-learning/