2017 Fifteenth International Conference on ICT and Knowledge Engineering

Artificial Intelligence, Machine Learning and Deep Learning

Pariwat Ongsulee
Department of Computer Science
Faculty of Science, Siam University
Bangkok, Thailand
pariwatsiam@gmail.com

***Abstract*—It is increasingly recognized that artificial intelligence has been touted as a new mobile. Because of the high volume of data that being generated by devices, sensors and social media users, the machine can learn to distinguish the pattern and makes a reasonably good prediction. This article will explore the use of machine learning and its methodologies. Furthermore, the field of deep learning which is being exploited in many leading IT providers will be clarified and discussed.**

***Keywords*—Artificial Intelligence, machine learning, deep learning, artificial neural network**

I. Introduction

Artificial intelligence (AI) is intelligence exhibited by machines. In computer science, the field of AI research defines itself as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of success at some goal. [1] Colloquially, the term "artificial intelligence" is applied when a machine mimics "cognitive" functions that humans associate with other human minds, such as "learning" and "problem solving". Capabilities currently classified as AI include successfully understanding human speech, [2] competing at a high level in strategic game systems (such as chess and Go [3]), self-driving cars, intelligent routing in content delivery networks, military simulations, and interpreting complex data.

The central problems (or goals) of AI research include reasoning, knowledge, planning, learning, natural language processing (communication), perception and the ability to move and manipulate objects.[4] General intelligence is among the field's long-term goals.[5] Approaches include statistical methods, computational intelligence, and traditional symbolic AI. Many tools are used in AI, including versions of search and mathematical optimization, logic, methods based on probability and economics. The AI field draws upon computer science, mathematics, psychology, linguistics, philosophy, neuroscience, artificial psychology plus many others.

AI has been part of our thoughts and slowly evolving in academic research labs since a group of computer scientists first defined the term at the Dartmouth Conferences in 1956 [6] and provided the genesis of the field. In the long decades since, AI has alternately been heralded as an all-encompassing Holy Grail and thrown into technology's bit bucket as a conception of overactive academic imaginations. In reality, however, until around 2012, its reach was limited to advanced technological companies, governments, and research agencies, feeding into both perceptions.

Since then, AI has broken away from the hypothetical and into real-world business solutions. Much of them have to do with the wide availability of GPUs (graphics processing units), which make parallel processing ever faster, cheaper, and more powerful. The ascent of AI also has to do with the simultaneous of practically infinite storage and a deluge of data of every stripe, including images, video, audio, text, transactions, and geospatial data.

II. Ease of Use

A. The Graphics Processing Units

A graphics processing unit (GPU), occasionally called visual processing unit (VPU), is a specialized electronic circuit designed to rapidly manipulate and alter memory to accelerate the creation of images in a frame buffer intended for output to a display device. GPUs are used in embedded systems, mobile phones, personal computers, workstations, and game consoles. Modern GPUs are very efficient at manipulating computer graphics and image processing, and their highly parallel structure makes them more efficient than general-purpose CPUs for algorithms where the processing of large blocks of data is done in parallel. In a personal computer, a GPU can be present on a video card, or it can be embedded on the motherboard or—in certain CPUs—on the CPU die. [7]

The term GPU was popularized by Nvidia in 1999, who marketed the GeForce 256 as "the world's first GPU", or Graphics Processing Unit. [8] It was presented as a "single-chip processor with integrated transform, lighting, triangle setup/clipping, and rendering engines". [9] Rival ATI Technologies coined the term "visual processing unit" or VPU with the release of the Radeon 9700 in 2002. [10]

Computer processors are designed to handle pretty much anything. Central Processing Units (CPUs), however, are very restricted and as such, can only perform certain mathematical calculations. Highly complicated combinations are not

978-1-5386-2117-2/17/$31.00 ©2017 IEEE

practical due to very long processing time. GPUs, on the other hand, have become so specialized that they surpass traditional processors when it comes to rendering large amounts of complex calculations.

Some examples include pedestrian detection for autonomous driving, medical imaging, supercomputing and machine learning. This comes as no surprise, because GPUs offer 10 to 100 times more computational power than traditional CPUs, which is one of the main reasons why graphics cards are currently being used to power some of the most advanced neural networks responsible for deep learning.

Some people might think that it has to do with parallelism, but the real reason is simpler than that and has to do with memory bandwidth. CPUs are capable of fetching small packages of memory quickly whereas GPUs have a high latency which makes them slower at this type of work. But GPUs are ideal when it comes to fetching very large amounts of memory and the best GPUs can fetch up to 750GB/s, which is huge when compare it to the best CPU which can handle only up to 50GB/s memory bandwidth.

To overcome the latency issues, we use more than one processing unit. GPUs are comprised of thousands of cores, unlike CPUs and to solve a task involving large amounts of memory and matrices, you would only have to wait for the initial fetch to take place. Every subsequent fetch will be significantly faster, due to the unloading process taking so much time that all the GPU have to queue in order to continue the unloading process. With so much processing power, the latency is effectively masked in order to allow the GPU to handle high bandwidth. This is called thread parallelism and it's the second reason why GPUs outperform traditional CPUs when it comes to deep learning.

The third reason is not that important performance-wise, but it does offer an additional insight into GPUs' undeniable supremacy over CPUs. The first part of the process involves fetching memory from the main or RAM memory and transferring it over to on-chip memory, or the L1 cache (instruction memory) and registers. Registers are attached directly to the execution unit, which for GPUs is the stream processor and for CPUs the core. This is where all the computation happens. Normally, you'd want both L1 and register memory to be as close to the execution engine and allow for a quick access by keeping the memories small. The larger the memory, the more time you need to access it.

In 2010, Nvidia began a partnership with Audi to power their cars' dashboards. These Tegra GPUs were powering the cars' dashboard, offering increased functionality to cars' navigation and entertainment systems.[11] Advancements in GPU technology in cars has helped push self-driving technology.[12] As of 2016 GPUs are popular for AI work, and they continue to evolve in a direction to facilitate deep learning, both for training[13] and inference in devices such as self-driving cars.[14] - and gaining additional connective capability for the kind of data-flow workloads AI benefits from (e.g. NVidia NVLink).[15]

B. *Machine Learning*

Machine learning is the subfield of computer science that, according to Arthur Samuel in 1959, gives "computers the ability to learn without being explicitly programmed."[16] Evolved from the study of pattern recognition and computational learning theory in artificial intelligence,[17]machine learning explores the study and construction of algorithms that can learn from and make predictions on data[18] – such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or unfeasible; example applications include email filtering, detection of network intruders or malicious insiders working towards a data breach,[19] optical character recognition (OCR),[20] learning to rank and computer vision.

Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining,[21] where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning.[22] Machine learning can also be unsupervised[23] and be used to learn and establish baseline behavioral profiles for various entities[24] and then used to find meaningful anomalies.

Within the field of data analytics, machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analytics. These analytical models allow researchers, data scientists, engineers, and analysts to "produce reliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data.

C. *The Methods of Machine Learning*[25]

Two of the most widely adopted machine learning methods are supervised learning and unsupervised learning. Most machine learning – about 70 percent – is supervised learning. Unsupervised learning accounts for 10 to 20 percent. Semi-supervised and reinforcement learning are two other technologies that are sometimes used.

• Supervised learning algorithms are trained using labeled examples, such as an input where the desired output is known. For example, a piece of equipment could have data points labeled either "F" (failed) or "R" (runs). The learning algorithm receives a set of inputs along with the corresponding correct outputs, and the algorithm learns by comparing its actual output with correct outputs to find errors. It then modifies the model accordingly. Through methods like classification, regression, prediction and gradient boosting, supervised learning uses patterns to predict the values of the label on additional unlabeled data. Supervised learning is commonly used in applications where historical data predicts likely future events. For example, it can anticipate when credit card transactions are likely to be fraudulent or which insurance customer is likely to file a claim.

• Unsupervised learning is used against data that has no historical labels. The system is not told the "right answer." The algorithm must figure out what is being shown. The goal is to explore the data and find some structure within. Unsupervised learning works well on transactional data. For example, it can identify segments of customers with similar attributes who can then be treated similarly in marketing campaigns. Or it can find the main attributes that separate customer segments from each other. Popular techniques include self-organizing maps, nearest-neighbor mapping, k-means clustering and singular value decomposition. These algorithms are also used to segment text topics, recommend items and identify data outliers.

•Semi-supervised learning is used for the same applications as supervised learning. But it uses both labeled and unlabeled data for training – typically a small amount of labeled data with a large amount of unlabeled data (because unlabeled data is less expensive and takes less effort to acquire). This type of learning can be used with methods such as classification, regression and prediction. Semi-supervised learning is useful when the cost associated with labeling is too high to allow for a fully labeled training process. Early examples of this include identifying a person's face on a web cam.

•Reinforcement learning is often used for robotics, gaming and navigation. With reinforcement learning, the algorithm discovers through trial and error which actions yield the greatest rewards. This type of learning has three primary components: the agent (the learner or decision maker), the environment (everything the agent interacts with) and actions (what the agent can do). The objective is for the agent to choose actions that maximize the expected reward over a given amount of time. The agent will reach the goal much faster by following a good policy. So the goal in reinforcement learning is to learn the best policy.

D. *Predictive Analytics*

Predictive analytics encompasses a variety of statistical techniques from predictive modeling, machine learning, and data mining that analyze current and historical facts to make predictions about future or otherwise unknown events.[26][27]

In business, predictive models exploit patterns found in historical and transactional data to identify risks and opportunities. Models capture relationships among many factors to allow assessment of risk or potential associated with a particular set of conditions, guiding decision making for candidate transactions. [28]

The defining functional effect of these technical approaches is that predictive analytics provides a predictive score (probability) for each individual (customer, employee, healthcare patient, product SKU, vehicle, component, machine, or other organizational unit) in order to determine, inform, or influence organizational processes that pertain across large numbers of individuals, such as in marketing, credit risk assessment, fraud detection, manufacturing, healthcare, and government operations including law enforcement.

Predictive analytics is used in actuarial science,[29] marketing,[30]financial services,[31] insurance, telecommunications,[32] retail,[33] travel,[34] mobility,[35] healthcare,[36] child protection,[37][38] pharmaceuticals,[39] capacity planning[40] and other fields.

One of the best-known applications is credit scoring, [26] which is used throughout financial services. Scoring models process a customer's credit history, loan application, customer data, etc., in order to rank-order individuals by their likelihood of making future credit payments on time.

E. *Deep Learning*

Deep learning (also known as deep structured learning, hierarchical learning or deep machine learning) is the study of artificial neural networks and related machine learning algorithms that contain more than one hidden layer. These deep nets: [41]

• use a cascade of many layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. The algorithms may be supervised or unsupervised and applications include pattern analysis (unsupervised) and classification (supervised).

• are based on the (unsupervised) learning of multiple levels of features or representations of the data. Higher level features are derived from lower level features to form a hierarchical representation.

• are part of the broader machine learning field of learning representations of data.

• learn multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts.

In a simple case, there might be two sets of neurons: one set that receives an input signal and one that sends an output signal. When the input layer receives an input it passes on a modified version of the input to the next layer. In a deep network, there are many layers between the input and the output (and the layers are not made of neurons but it can help to think of it that way), allowing the algorithm to use multiple processing layers, composed of multiple linear and non-linear transformations [42]

Deep learning is part of a broader family of machine learning methods based on learning representations of data. An observation (e.g., an image) can be represented in many ways such as a vector of intensity values per pixel, or in a more abstract way as a set of edges, regions of particular shape, etc. Some representations are better than others at simplifying the learning task (e.g., face recognition or facial expression recognition [43]). One of the promises of deep learning is replacing handcrafted features with efficient algorithms for unsupervised or semi-supervised feature learning and hierarchical feature extraction. [44]

Research in this area attempts to make better representations and create models to learn these representations from large-scale unlabeled data. Some of the representations are inspired by advances in neuroscience and are loosely based on interpretation of information processing and communication patterns in a nervous system, such as neural coding which attempts to define a relationship between various stimuli and associated neuronal responses in the brain.[45]

Various deep learning architectures such as deep neural networks, convolutional deep neural networks, deep belief networks and recurrent neural networks have been applied to fields like computer vision, automatic speech recognition, natural language processing, audio recognition and bioinformatics where they have been shown to produce state-of-the-art results on various tasks.

Although Deep learning has been characterized as a buzzword, or a rebranding of neural networks[46][47], deep neural nets have demonstrated an ability to out-perform other machine learning algorithms on tasks such as object recognition in the field of computer vision.[48]

F. *Artificial Neural Networks*

Some of the most successful deep learning methods involve artificial neural networks. Artificial neural networks are inspired by the 1959 biological model proposed by Nobel laureates David H. Hubel & Torsten Wiesel, who found two types of cells in the primary visual cortex: simple cells and complex cells. Many artificial neural networks can be viewed as cascading models [49] of cell types inspired by these biological observations.

Fukushima's Neocognitron introduced convolutional neural networks partially trained by unsupervised learning with human-directed features in the neural plane. Yann LeCun et al. (1989) applied supervised backpropagation to such architectures.[50] Weng et al. (1992) published convolutional neural networks Cresceptron[51] for 3-D object recognition from images of cluttered scenes and segmentation of such objects from images.

The use of the expression "Deep Learning" in the context of Artificial Neural Networks was introduced by Igor Aizenberg and colleagues in 2000.[52] A Google Ngram chart shows that the usage of the term has gained traction (actually has taken off) since 2000.[53] In 2006, a publication by Geoffrey Hinton and Ruslan Salakhutdinov drew additional attention by showing how many-layered feedforward neural network could be effectively pre-trained one layer at a time, treating each layer in turn as an unsupervised restricted Boltzmann machine, then fine-tuning it using supervised backpropagation.[54] In 1992, Schmidhuber had already implemented a very similar idea for the more general case of unsupervised deep hierarchies of recurrent neural networks, and also experimentally shown its benefits for speeding up supervised learning.

In 2012, the Google Brain team led by Andrew Ng and Jeff Dean created a neural network that learned to recognize higher-level concepts, such as cats, only from watching unlabeled images taken from YouTube videos.[55][56]

G. *Convolutional Neural Networks*

In machine learning, a convolutional neural network (CNN, or ConvNet) is a type of feed-forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex. Individual cortical neurons respond to stimuli in a restricted region of space known as the receptive field. The receptive fields of different neurons partially overlap such that they tile the visual field. The response of an individual neuron to stimuli within its receptive field can be approximated mathematically by a convolution operation.[57] Convolutional networks were inspired by biological processes[58] and are variations of multilayer perceptrons designed to use minimal amounts of preprocessing.[59] They have wide applications in image and video recognition, recommender systems[60] and natural language processing.[61]

For example, in the image recognition application, convolutional neural networks (CNNs) consist of multiple layers of receptive fields. These are small neuron collections which process portions of the input image. The outputs of these collections are then tiled so that their input regions overlap, to obtain a higher-resolution representation of the original image; this is repeated for every such layer. Tiling allows CNNs to tolerate translation of the input image.[62]

Convolutional networks may include local or global pooling layers, which combine the outputs of neuron clusters.[63][64] They also consist of various combinations of convolutional and fully connected layers, with pointwise nonlinearity applied at the end of or after each layer.[65] A convolution operation on small regions of input is introduced to reduce the number of free parameters and improve generalization. One major advantage of networks is the use of shared weight in convolutional layers, which means that the same filter (weights bank) is used for each pixel in the layer; this both reduces memory footprint and improves performance

III. CONCLUSION

Artificial Intelligence, machine learning and deep learning are basically machine perception. It is the power to interpret sensory data. Two main ways we interpret things are by naming what we sense; e.g. we hear a sound as we say ourselves "That's my daughter's voice." Or we see a haze of photons and we say "That's my mother's face." If we don't have names for things, we can still recognize similarities and dissimilarities. You might see two faces and know that they were mother and daughter, without knowing their names; or you might hear two voices and know that they came from the same town or state by their accent. Algorithms train to name things through supervised learning, and to cluster things through unsupervised learning. The difference between supervised and unsupervised learning is whether you have a labeled training set to work with or not. The labels you apply to data are simply the outcomes you care about. Maybe you care about identifying people in images. Maybe you care about identifying angry or spammy emails, which are all just unstructured blobs of text. Maybe you're looking at time series data -- a stream of numbers -- and you care about whether the next instances in the time series will be higher or lower.

So deep learning, working with other algorithms, can help to classify, cluster and predict. It does so by learning to read the signals, or structure, in data automatically. When deep learning algorithms train, they make guesses about the data, measure the error of their guesses against the training set, and then correct the way they make guesses in order to become more accurate. This is optimization.

Now imagine that, with deep learning, you can classify, cluster or predict anything you have data about: images, video, sound, text and DNA, time series (touch, stock markets, economic tables, and the weather). That is, anything that humans can sense and that our technology can digitize. You have multiplied your ability to analyze what's happening in the world by many times. With deep learning, we are basically giving society the ability to behave much more intelligently, by accurately interpreting what's happening in the world around us with software.

Prediction alone is a huge power, and the applications are fairly obvious. Classification sounds banal, but by naming something, you can decide how to respond. If an email is spam, you send it to the spam folder and save the reader time. If the face captured by your front door camera is your mother, maybe you tell the smart lock to open the door. If an X-ray shows a tumorous pattern, you flag it for deeper examination by medical experts.

References

[1] "The intelligent agent paradigm", Russell & Norvig 2003, pp. 27, 32–58, 968–972, Poole, Mackworth & Goebel 1998, pp. 7–21, Luger & Stubblefield 2004, pp. 235–240, Hutter 2005, pp. 125–126

[2] Russell & Norvig 2009.

[3] "AlphaGo - Google DeepMind".

[4] This list of intelligent traits is based on the topics covered by the major AI textbooks, including:Russell & Norvig 2003, Luger & Stubblefield 2004, Poole, Mackworth & Goebel 1998, Nilsson 1998.

[5] General intelligence (strong AI) is discussed in popular introductions to AI: Kurzweil 1999 and Kurzweil 2005

[6] Dartmouth conference: McCorduck 2004, pp. 111–136, Crevier 1993, pp. 47–49, who writes "the conference is generally recognized as the official birthdate of the new science.", Russell & Norvig 2003, p. 17, who call the conference "the birth of artificial intelligence.", NRC 1999, pp. 200–201

[7] Denny Atkin. "Computer Shopper: The Right GPU for You". Retrieved 2007-05-15.

[8] "NVIDIA Launches the World's First Graphics Processing Unit: GeForce 256". Nvidia. 31 August 1999. Retrieved 28 March 2016 from http://www.nvidia.com/object/IO_20020111_5424.html

[9] "Graphics Processing Unit (GPU)". Nvidia. Retrieved 29 March 2016 from http://www.nvidia.com/object/gpu.html

[10] Pabst, Thomas (18 July 2002). "ATi Takes Over 3D Technology Leadership With Radeon 9700". Tom's Hardware. Retrieved 29 March 2016 from http://www.tomshardware.com/reviews/ati-takes-3d-technology-leadership-radeon-9700,491-3.html

[11] Teglet, Traian. "NVIDIA Tegra Inside Every Audi 2010 Vehicle". Retrieved 2016-08-03 from http://news.softpedia.com/news/NVIDIA-Tegra-Inside-Every-Audi-2010-Vehicle-131529.shtml

[12] "School's in session — Nvidia's driverless system learns by watching". 2016-04-30. Retrieved 2016-08-03 from https://www.digitaltrends.com/cars/nvidia-gpu-driverless-car/

[13] "Nvidia driving the development of deep learning". 2016-05-17 Retrieved 2017-06-01 from https://insidehpc.com/2016/05/nvidia-driving-the-development-of-deep-learning/

[14] Hanley, Steve "Nvidia introduces supercomputer for self driving cars". 2016-1-6 Retrieved 2017-06-01 from http://gas2.org/2016/01/06/nvidia-introduces-supercomputer-for-self-driving-cars/

[15] Harris, Mark How NVLink Will Enable Faster, Easier Multi-GPU Computing. 2014-11-14 Retrieved 2017-06-01 from https://devblogs.nvidia.com/parallelforall/how-nvlink-will-enable-faster-easier-multi-gpu-computing/

[16] Munoz, Andres, Machine Learning and Optimization [Online] Retrieved 2017-06-01 from https://www.cims.nyu.edu/~munoz/files/ml_optimization.pdf

[17] Hosch, William L., Machine Learning [Online] Retrieved 2017-06-01 from http://www.britannica.com/EBchecked/topic/1116194/machine-learning

[18] Ron Kohavi; Foster Provost (1998). "Glossary of terms". Machine Learning. 30: 271–274.

[19] Dickson, Ben. "Exploiting machine learning in cybersecurity". TechCrunch. Retrieved 2017-05-23. from https://techcrunch.com/2016/07/01/exploiting-machine-learning-in-cybersecurity/

[20] Wernick, Yang, Brankov, Yourganov and Strother, Machine Learning in Medical Imaging, IEEE Signal Processing Magazine, vol. 27, no. 4, July 2010, pp. 25–38

[21] Mannila, Heikki (1996). Data mining: machine learning, statistics, and databases. Int'l Conf. Scientific and Statistical Database Management. IEEE Computer Society.

[22] Friedman, Jerome H. (1998). "Data Mining and Statistics: What's the connection?". Computing Science and Statistics. 29 (1): 3–9.

[23] Engel, Giora, 3 Flavors of Machine Learning: Who, What & Where. DarkReading, 2016-11-2 from http://www.darkreading.com/threat-intelligence/3-flavors-of-machine-learning--who-what-and-where/a/d-id/1324278

[24] LightCyber's Jason Matlof Explains how Magna Detects Hackers Before They Attack. AI Business, 2016-10-14 from http://aibusiness.org/lightcybers-jason-matlof-explains-how-magna-detects-hackers-before-they-attack/#sthash.aIGRF1An.dpuf

[25] "Machine Learning: What it is and why it matters". www.sas.com. Retrieved 2016-03-29 from https://www.sas.com/it_it/insights/analytics/machine-learning.html

[26] Nyce, Charles (2007), Predictive Analytics White Paper (PDF), American Institute for Chartered Property Casualty Underwriters/Insurance Institute of America, p. 1

[27] Eckerson, Wayne (May 10, 2007), Extending the Value of Your Data Warehousing Investment, The Data Warehouse Institute

[28] Coker, Frank (2014). Pulse: Understanding the Vital Signs of Your Business (1st ed.). Bellevue, WA: Ambient Light Publishing. pp. 30, 39, 42,more. ISBN 978-0-9893086-0-1.

[29] Conz, Nathan (September 2, 2008), "Insurers Shift to Customer-focused Predictive Analytics Technologies", Insurance & Technology from http://www.insurancetech.com/data-and-analytics/insurers-shift-to-customer-focused-predictive-analytics-technologies/d/d-id/1310535?

[30] Fletcher, Heather (March 2, 2011), "The 7 Best Uses for Predictive Analytics in Multichannel Marketing", Target Marketing from http://www.targetmarketingmag.com/article/7-best-uses-predictive-analytics-modeling-multichannel-marketing/#

[31] Korn, Sue (April 21, 2011), "The Opportunity for Predictive Analytics in Finance", HPC Wire from https://www.hpcwire.com/hpcwire/2011-04-21/the_opportunity_for_predictive_analytics_in_finance.html

[32] Barkin, Eric (May 2011), "CRM + Predictive Analytics: Why It All Adds Up", Destination CRM from http://www.destinationcrm.com/Articles/Editorial/Magazine-Features/CRM---Predictive-Analytics-Why-It-All-Adds-Up-74700.aspx

[33] Das, Krantik; Vidyashankar, G.S. (July 1, 2006), "Competitive Advantage in Retail Through Analytics: Developing Insights, Creating

Value", Information Management from https://www.information-management.com/infodirect/20060707/1057744-1.html

[34] McDonald, Michèle (September 2, 2010), "New Technology Taps 'Predictive Analytics' to Target Travel Recommendations", Travel Market Report

[35] Moreira-Matias, Luís; Gama, João; Ferreira, Michel; Mendes-Moreira, João; Damas, Luis (2016-02-01). "Time-evolving O-D matrix estimation using high-speed GPS data streams". Expert Systems with Applications. 44: 275–288.

[36] Stevenson, Erin (December 16, 2011), "Tech Beat: Can you pronounce health care predictive analytics?", Times-Standard

[37] Lindert, Bryan (October 2014). "Eckerd Rapid Safety Feedback Bringing Business Intelligence to Child Welfare" (PDF). Policy & Practice. Retrieved March 3, 2016. from http://static.eckerd.org/wp-content/uploads/Eckerd.pdf

[38] "Florida Leverages Predictive Analytics to Prevent Child Fatalities -- Other States Follow". The Huffington Post. Retrieved 2016-03-25. from http://www.huffingtonpost.com/marquis-cabrera/florida-leverages-predictive_b_8586712.html

[39] McKay, Lauren (August 2009), "The New Prescription for Pharma", Destination CRM from http://www.destinationcrm.com/articles/Web-Exclusives/Web-Only-Bonus-Articles/The-New-Prescription-for-Pharma-55774.aspx

[40] Diaz, Julian (May 2016) Using Predictive Analytics for effective Capacity Management, Principa from http://insights.principa.co.za/using-predictive-analytics-for-effective-capacity-management

[41] Deng, L.; Yu, D. (2014). "Deep Learning: Methods and Applications" (PDF). Foundations and Trends in Signal Processing. 7 (3–4): 199–200. doi:10.1561/2000000039.

[42] Ian Goodfellow, Yoshua Bengio, and Aaron Courville (2016). Deep Learning. MIT Press. Online from http://www.deeplearningbook.org

[43] Glauner, P. (2015). Deep Convolutional Neural Networks for Smile Recognition (MSc Thesis). Imperial College London, Department of Computing. arXiv:1508.06535

[44] Song, H.A.; Lee, S. Y. (2013). "Hierarchical Representation Using NMF". Neural Information Processing. Lectures Notes in Computer Sciences. 8226. Springer Berlin Heidelberg. pp. 466–473. doi:10.1007/978-3-642-42054-2_58. ISBN 978-3-642-42053-5.

[45] Olshausen, B. A. (1996). "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". Nature. 381 (6583): 607–609. doi:10.1038/381607a0. PMID 8637596 from http://www.nature.com/nature/journal/v381/n6583/abs/381607a0.html

[46] Collobert, R. (April 2011). Deep Learning for Efficient Discriminative Parsing. VideoLectures.net. Event occurs at 7min 45s. from http://videolectures.net/aistats2011_collobert_deep/

[47] Gomes, L. (20 October 2014). "Machine-Learning Maestro Michael Jordan on the Delusions of Big Data and Other Huge Engineering Efforts". IEEE Spectrum. from http://spectrum.ieee.org/robotics/artificial-intelligence/machinelearning-maestro-michael-jordan-on-the-delusions-of-big-data-and-other-huge-engineering-efforts

[48] Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffry (2012). "ImageNet Classification with Deep Convolutional Neural Networks" (PDF). NIPS 2012: Neural Information Processing Systems, Lake Tahoe, Nevada.

[49] J. Weng, N. Ahuja and T. S. Huang, "Cresceptron: a self-organizing neural network which grows adaptively," Proc. International Joint Conference on Neural Networks, Baltimore, Maryland, vol I, pp. 576-581, June, 1992.

[50] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel. 1989 Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation, 1(4):541–551.

[51] J. Weng, N. Ahuja and T. S. Huang, "Learning recognition and segmentation of 3-D objects from 2-D images," Proc. 4th International Conf. Computer Vision, Berlin, Germany, pp. 121-128, May, 1993.

[52] Igor Aizenberg, Naum N. Aizenberg, Joos P.L. Vandewalle (2000). Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications. Springer Science & Business Media.

[53] Google Ngram chart of the usage of the expression "deep learning" posted by Jürgen Schmidhuber (2015) from https://plus.google.com/100849856540000067209/posts/7N6z251w2Wd?pid=6127540521703625346&oid=100849856540000067209

[54] G. E. Hinton., "Learning multiple layers of representation," Trends in Cognitive Sciences, 11, pp. 428–434, 2007

[55] John Markoff (25 June 2012). "How Many Computers to Identify a Cat? 16,000.". New York Times.

[56] Ng, Andrew; Dean, Jeff (2012). "Building High-level Features Using Large Scale Unsupervised Learning". arXiv:1112.6209

[57] "Convolutional Neural Networks (LeNet) – DeepLearning 0.1 documentation". DeepLearning 0.1. LISA Lab. Retrieved 31 August 2013. from http://deeplearning.net/tutorial/lenet.html

[58] Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda (2003). "Subject independent facial expression recognition with robust face detection using a convolutional neural network" (PDF). Neural Networks. 16 (5): 555–559. doi:10.1016/S0893-6080(03)00115-1. Retrieved 17 November 2013. from http://www.iro.umontreal.ca/~pift6080/H09/documents/papers/sparse/matsugo_etal_face_expression_conv_nnet.pdf

[59] LeCun, Yann. "LeNet-5, convolutional neural networks". Retrieved 16 November 2013. from http://yann.lecun.com/exdb/lenet/

[60] van den Oord, Aaron; Dieleman, Sander; Schrauwen, Benjamin (2013-01-01). Burges, C. J. C.; Bottou, L.; Welling, M.; Ghahramani, Z.; Weinberger, K. Q., eds. Deep content-based music recommendation (PDF). Curran Associates, Inc. pp. 2643–2651.

[61] Collobert, Ronan; Weston, Jason (2008-01-01). "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning". Proceedings of the 25th International Conference on Machine Learning. ICML '08. New York, NY, USA: ACM: 160–167. ISBN 978-1-60558-205-4. doi:10.1145/1390156.1390177.

[62] Korekado, Keisuke; Morie, Takashi; Nomura, Osamu; Ando, Hiroshi; Nakano, Teppei; Matsugu, Masakazu; Iwata, Atsushi (2003). "A Convolutional Neural Network VLSI for Image Recognition Using Merged/Mixed Analog-Digital Architecture". Knowledge-Based Intelligent Information and Engineering Systems: 169–176. CiteSeerX 10.1.1.125.3812

[63] Ciresan, Dan; Ueli Meier; Jonathan Masci; Luca M. Gambardella; Jurgen Schmidhuber (2011). "Flexible, High Performance Convolutional Neural Networks for Image Classification" (PDF). Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Two. 2: 1237–1242. Retrieved 17 November 2013. from http://people.idsia.ch/~juergen/ijcai2011.pdf

[64] Krizhevsky, Alex. "ImageNet Classification with Deep Convolutional Neural Networks" (PDF). Retrieved 17 November 2013. from http://www.image-net.org/challenges/LSVRC/2012/supervision.pdf

[65] Ciresan, Dan; Meier, Ueli; Schmidhuber, Jürgen (June 2012). "Multi-column deep neural networks for image classification". 2012 IEEE Conference on Computer Vision and Pattern Recognition. New York, NY: Institute of Electrical and Electronics Engineers (IEEE): 3642–3649. ISBN 978-1-4673-1226-4. OCLC 812295155.arXiv:1202.2745v1