Artificial Neural Network Modelling: An Introduction

Subana Shanmuganathan

Abstract While scientists from different disciplines, such as neuroscience, medicine and high performance computing, eagerly attempt to understand how the human brain functioning happens, Knowledge Engineers in computing have been successful in making use of the brain models thus far discovered to introduce heuristics into computational algorithmic modelling. Gaining further understanding on human brain/nerve cell anatomy, structure, and how the human brain functions, is described to be significant especially, to devise treatments for presently described as incurable brain and nervous system related diseases, such as Alzheimer's and epilepsy. Despite some major breakthroughs seen over the last few decades neuroanatomists and neurobiologists of the medical world are yet to understand how we humans think, learn and remember, and how our cognition and behaviour are linked. In this context, the chapter outlines the most recent human brain research initiatives following which early Artificial Neural Network (ANN) architectures, components, related terms and hybrids are elaborated.

1 Introduction

Neuroanatomists and Neurobiologists of the medical world are yet to discover the exact structure and the real processing that takes place in human nerve cells and to biologically model the human brain. This is despite the breakthroughs made by research ever since the human beings themselves began wondering how their own thinking ability happens. More recently, there has been some major initiatives with unprecedented funding, that emphasise the drive, to accelerate research into unlocking the mysteries of human brain's unique functioning. One among such big funding projects is the Human Brain Project (HBP) initiated in 2013. The HBP is a European Commission Future and Emerging Technologies Flagship that aims to

S. Shanmuganathan (✉)
Auckland University of Technology, Auckland, New Zealand
e-mail: Subana@gmail.com

understand what makes the brain unique, the basic mechanisms behind cognition and behaviour, how to objectively diagnose brain diseases, and to build new technologies inspired by how the brain computes. There are 13 subprojects (SPs) within this ten-year one-billion pound HBP programme. The scientists involved in the HBP accept that the current computer technology is insufficient to simulate complex brain functioning. However, they are hopeful of having sufficiently powerful supercomputers to begin the first draft simulation of the human brain within a decade. It is surprising that despite the remarkable and ground breaking innovations achieved in computing leading to transformations never seen in human development, even the modern day's most powerful computers still struggle to do things that humans find instinctive. "Even very young babies can recognise their mothers but programming a computer to recognise a particular person is possible but very hard." [1]. Hence, SP9 scientists of HBP are working on developing "neuromorphic computers-machines" that can learn in a similar manner to how the brain functions. The other major impediment in this regard is the humongous amount of data that will be produced, which is anticipated to require massive amount of computing memory. Currently, HBP scientists of The SpiNNaker project at the University of Manchester are building a model, which will mimic 1 % of brain function. Unlocking brain functioning secrets in this manner is anticipated to yield major benefits in information technology as well. The advent of neuromorphic computers and knowledge could lead to the production of computer chips with specialised cognitive skills that truly mimic those of the human brain, such as the ability to analyse crowds, or decision-making on large and complex datasets. These digital brains should also allow researchers to compare healthy and diseased brains within computer models [2].

Meanwhile, across the Atlantic, the unveiling of Brain Research Through Advancing Innovative Neurotechnologies—or BRAIN in the USA by President Obama took place in 2013 [3]. This was announced to keep up with the brain research initiated in Europe. The BRAIN project was said to begin in 2014 and be carried out by both public and private-sector scientists to map the human brain. The President announced an initial $100 m investment to shed light on how the brain works and to provide insight into diseases such as Alzheimer's, Parkinson's, epilepsy and many more. At the White House inauguration, President Obama said: "There is this enormous mystery waiting to be unlocked, and the BRAIN initiative will change that by giving scientists the tools they need to get a dynamic picture of the brain in action and to better understand how we think and learn and remember. And that knowledge will be transformative." In addition, the US President as well pointed out a lack of research in this regard, "As humans we can identify galaxies light years away, we can study particles smaller than the atom, but we still haven't unlocked the mystery of the 3 lb of matter that sits between our ears," [3].

With that introduction to contemporary research initiatives to unlock unique human brain functioning, Sect. 2 looks at the early brain models in knowledge engineering following which initial ANN models and their architectures are elaborated. In the final section some modern day ANN hybrids are outlined.

2 Early Brain Models in Knowledge Engineering

Using the Brain models developed based on our understanding thus far made on human like thinking, researchers in "Knowledge Engineering" continue to introduce functional models simulating the heuristic ability that is still considered as a unique characteristics of human intelligence. "We [Knowledge Engineers] are surprisingly flexible in processing information in the real world…" [4]. As reiterated by HRP as well as the US brain research initiatives of this decade, the discovery of actual processing in the human brain (consisting of 10^{11} neurons, participating in perhaps 10^{15} interconnections over transmission paths) seems to be very unlikely to be made in the near future. Nevertheless, the functional models of the knowledge engineers and combinations of these models have been put into successful use in knowledge representation and processing, and are known as Artificial Intelligence (AI) in computing. In the last few decades, there has been considerable research carried out with an appreciable amount of success in using knowledge-based systems for solving problems those needed heuristics.

Brain functions are mainly processed in the form of algorithms suggested John Holland (1975) [5], who was the first to compare the heuristic methods for problem solving with nature's evolution process using genetic approaches and genetic algorithms. Genetic algorithms solve complex combinational and organizational problems with many variants i.e., genes, chromosomes, population, mutation. In [6], it is explained that the brain is capable of acquiring information-processing algorithms automatically. This kind of elucidation not only forms a basis for understanding the growth of the brain and the factors needed for mental growth, but also enables us to develop novel information processing methods. Expert systems are an instance of rule-based expressions of knowledge, represented in the conditional mathematical form of "if and then" causal relationships.

In the last few decades, the performance of conventional computing has been growing spectacularly [1]. The reasons for this have been; the falling cost of large data storage devices, the increasing ease of collecting data over networks, the development of robust and efficient machine learning algorithms to process this data along with the falling cost of computational power. They have indeed enabled the use of computationally intensive methods for data analysis [7]. The field of "data mining" also called as "knowledge discovery" is one among them that has already produced practical applications in many areas. i.e., analysing medical outcomes, predicting customer purchase behaviour, predicting the personal interests of Web users, optimising manufacturing process, predicting trends in stock markets, financial analysis and sales in real estate investment appraisal of land properties, most of them using past observational data.

"Traditionally, human experts have derived their knowledge that is described as explicit from their own personal observations and experience. With advancing computer technology, automated knowledge discovery has become an important AI research topic, as well as practical business application in an increasing number of organisations…" [8]. Knowledge discovery has been identified as a method of

learning the implicit knowledge that is defined as previously unknown, non-trivial knowledge hidden in the past data or observations.

Above all, our expectations from computers have been growing. "In 40 years' time people will be used to using conscious computers and you wouldn't buy a computer unless it was conscious…." [9] as envisioned by Aleksander in 1999.

3 Artificial Neural Networks and Their Components

An Artificial Neural Network (ANN) in simple terms is a biologically inspired computational model, which consists of processing elements (called neurons), and connections between them with coefficients (weights) bound to the connections. These connections constitute the neuronal structure and attached to this structure are training and recall algorithms. Neural networks are called the connectionist models because of the connections found between the neurons [10].

Deboeck and Kohonen [11] described Neural networks (NNs) as a collection of mathematical techniques that can be used for signal processing, forecasting and clustering and termed it as non-linear, multi-layered, parallel regression techniques. It is further stated that neural network modelling is like fitting a line, plane or hyper plane through a set of data points. A line, plane or hyper plane can be fitted through any data set to define the relationships that may exist between (what the user chooses to be) the inputs and the outputs; or it can be fitted for identifying a representation of the data on a smaller scale.

The first definition describes the ANN from its similarities to the human brain like functioning (Fig. 1) and the latter (Kohonen) in an application perspective.

It is truly accepted inclusive of recent brain research initiatives that the human brain is much more complicated as many of its cognitive functions are still unknown. However, the following are the main characteristics considered and described as common functions in real and artificial networks:
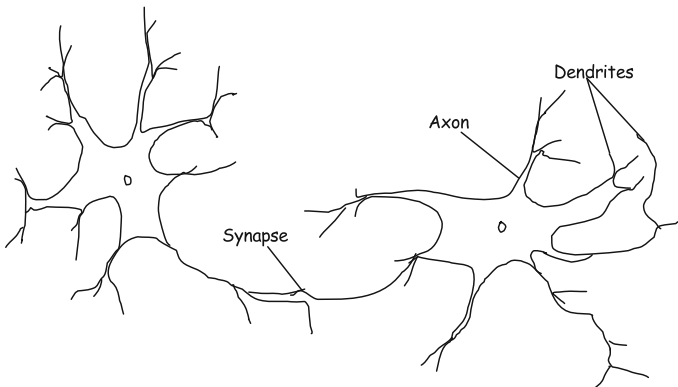

Fig. 1 Biological neuron

1. Learning and adaptation
2. Generalisation
3. Massive parallelism
4. Robustness
5. Associative storage of information
6. Spatiotemporal information processing.

Intrigued by the potentials of the ANNs, professionals from almost all fields are finding methods by way of creating new models using all possible combinations of symbolic and sub-symbolic paradigms, many of them with Fuzzy techniques to suit a variety of applications within their own disciplines.

McCulloch and Pitts were the first to introduce a mathematical model of a neuron in 1943. They continued their work [12] and explored network paradigms for pattern recognition despite rotation angle, translation, and scale factor related issues. Most of their work involved simple neuron model and these network systems were generally referred to as perceptrons.

The perceptron model of McCulloch and Pitts [12] created using Pitts and McCulloch [13] neuron is presented in Fig. 2. The Σ unit multiplies each input x by a weight w, and sums the weighted inputs. If this sum is greater than a predefined threshold, the output is one, otherwise zero. In general, they consist of a single layer. In 1958, using this neuron model of Pitts and McCulloch [13], Rosenblatt made a network with the aim of modelling the visual perception phenomena. In the 1960s, these perceptrons created a great interest and in 1962, Rosenblatt proved a theorem about perceptron learning. He showed that a perceptron could learn anything that it could represent. Consequently Widrow [14, 15], Widrow and Angell [16], and Widrow and Hoff [17] demonstrated convincing models. The whole world was exploring the potential of these perceptrons. But eventually as these single layer systems found to fail at certain simple learning tasks, researchers lost interest in ANNs. Consequently, Minsky [18] proved that single layer perceptrons had severe restrictions on their ability to represent and learn [18]. He further doubted a learning algorithm could be found for multi-layer neural networks. This caused almost an eclipse to ANN research and in turn made the researchers to develop symbolic AI methods and systems i.e., Expert Systems.

Later, from 1977 onwards, new connectionist models were introduced. Such as associative memories [19, 20], multi-layer perceptron (MLP) and back propagation learning algorithm [21, 22]; adaptive resonance theory (ART) [23, 24],
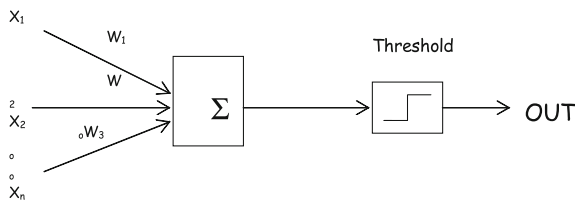

Fig. 2 Perceptron neuron of Pitts and McCulloch [12]

self-organising networks [25] and more. These new connectionist models drew the interest of many more researchers into sub-symbolic systems and as a result many more networks have been designed and used since then: i.e., Bi-directional asso-ciative memory introduced by [26], radial basis function by [27], probabilistic RAM neural networks of [28, 29], fuzzy neurons and fuzzy networks presented by [30, 31], oscillatory neurons and oscillatory neural networks [32–34] and many more. Based on these different neuron and network models enormous applications have been developed and successfully used in invariably all disciplines.

ANNs are increasingly being used across a variety of application areas where imprecise data or complex attribute relationships exist that are difficult to quantify using traditional analytical methods [10]. Research elaborated in the following chapters of this book show the more recent trends in ANN applications and the success achieved in using them.

The following are the parameters that describe a neuron based on Fig. 3.

1 Input connections (inputs): $x1$, $x2$, …, x_n. There are weights bound to the input connections: $w1$, $w2$, …, w_n. One input to the neuron, called the bias has a constant value of 1 and is usually represented as a separate input, let's refer to as $x0$, but for simplicity it is treated here just as an input, clamped to a constant value.
2 Input functions f: Calculates the aggregated net input signal to the neuron $u = f(x, w)$,
   where x and w are the input and weight vectors correspondingly;
   f is usually the summation function; $u = \Sigma i = 1, n x_i . w_i$
3 An activation (signal) function s calculates the activation level of the neuron

$$a = s(u).$$

4 An output function calculates the output signal value emitted through the output (the axon) of the neuron; $o = g(a)$; the output signal is usually assumed to be equal to the activation level of the neuron, that is, o = a.
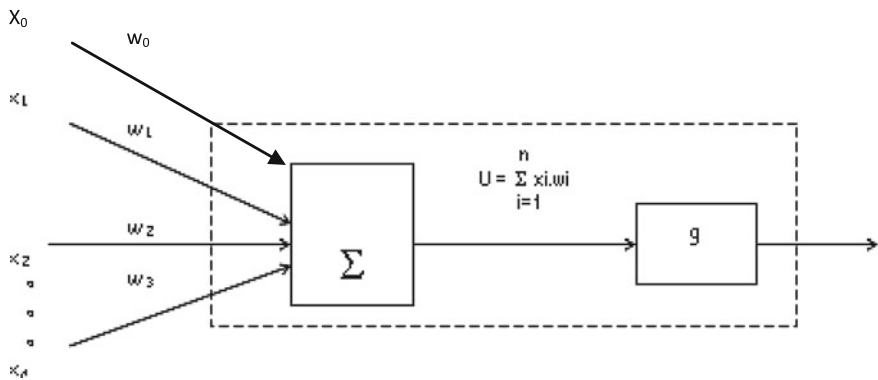

Fig. 3 A model of an artificial neuron

Artificial neural networks are usually defined by the following four parameters:

1 Type of neuron (or nodes as the neural network resembles a graph) i.e., Perceptron Pitts and McCulloch [13], Fuzzy neuron Yamakawa [30]
2 Connectionist architecture: The organisation of the connections between the neurons is described as the architecture. The connections between the neurons define the topology of the ANN. i.e., fully connected, partially connected (Fig. 4).
  Connectionist architecture can also be distinguished depending on the number of input and output neurons and the layers of neurons used

  (a) Autoassociative: Input neurons are the output neurons i.e., Hopfield network
  (b) Heteroassociative: There are separate input neurons and output neurons i.e., Multi-layer perceptron (MLP), Kohonen network.
      Furthermore, depending on the connections back from the output to the input neurons, two different kinds of architectures are determined:
  (a) Feedforward architecture: There are no connections back from the output neurons to the input neurons. The network does not remember of its previous output values and the activation states of its neurons.
  (b) Feedback architecture: There are connections back from the output neurons to the input neurons and as such the network holds in memory of its previous states and the next state depends on current input signals and the previous states of the network. i.e., Hopfield network.

3 Learning algorithm: is the algorithm, which trains the networks. Lots of research have been carried out in trying various phenomena and it gives the researchers an enormous amount of flexibility and opportunity for innovation and discussing the whole set of Learning algorithms is far beyond the scope of this chapter.
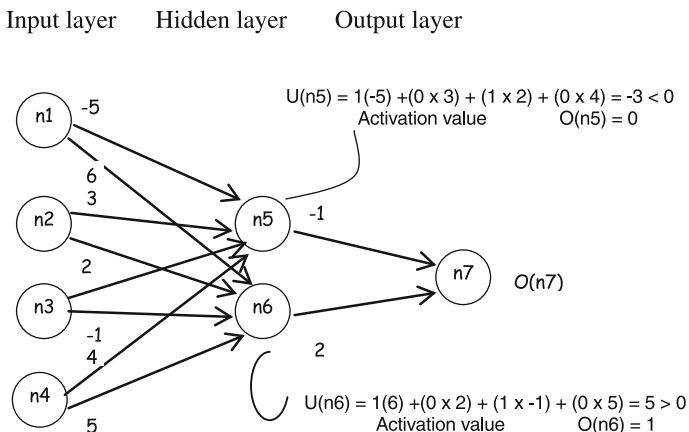

Input layer   Hidden layer   Output layer

$U(n5) = 1(-5) + (0 \times 3) + (1 \times 2) + (0 \times 4) = -3 < 0$
Activation value $O(n5) = 0$

$U(n6) = 1(6) + (0 \times 2) + (1 \times -1) + (0 \times 5) = 5 > 0$
Activation value $O(n6) = 1$

Fig. 4 A simple neural network with four input nodes (with an input vector 1, 0, 1, 0), two intermediate, and one output node. The connection weights are shown, presumably as a result of training

Nevertheless, the learning algorithms so far used are currently classified into three groups.

(a) Supervised learning: The training examples consist of input vectors x and the desired output vectors y and training is performed until the neural network "learns" to associate each input vector x to its corresponding output vector y (approximate a function $y = f(x)$). It encodes the example in its internal structure.

(b) Unsupervised learning: Only input vectors x are supplied and the neural network learns some internal feature of the whole set of all the input vectors presented to it. Contemporary unsupervised algorithms are further divided into two (i) noncompetitive and (ii) competitive.

(c) Reinforcement learning: Also referred to as reward penalty learning. The input vector is presented and the neural network is allowed to calculate the corresponding output and if it is good then the existing connection weights are increased (rewarded), otherwise the connection weights involved are decreased (punished).

4 Recall algorithm: By which learned knowledge is extracted from the network.

The following are the contemporary applications of ANN in general:

1 Function approximation, when a set of data is presented.
2 Pattern association.
3 Data clustering, categorisation, and conceptualisation.
4 Learning statistical parameters.
5 Accumulating knowledge through training.
6 "Extracting" knowledge through analysis of the connection weights.
7 Inserting knowledge in a neural network structure for the purpose of approximate reasoning.

The problem solving process using the neural networks actually consists of two major phases and they are:

(i) Training phase: During this phase the network is trained with training examples and the rules are inserted in its structure.
(ii) Recall phase: When new data is fed to the trained network the recall algorithm is used to calculate the results.

The problem solving process is described as mapping of problem domain, problem knowledge and solution space into the network's input state space, synaptic weights space and the output space respectively. Based on recent studies construction of a neural network could be broken into the following steps [10]:

(1) Problem identification: What is the generic problem and what kind of knowledge is available?
(2) Choosing an appropriate neural network model for solving the problem.
(3) Preparing data for training the network, which process may include statistical analysis, discretisation, and normalisation.

(4) Training a network, if data for training is available. This step may include creating a learning environment in which neural networks are "pupils".

(5) Testing the generalisation ability of the trained neural network and validating the results.

In recent years, neural networks have been considered to be universal function approximators. They are model free estimators [35]. Without knowing the type of the function, it is possible to approximate the function. However, the difficult part of it is, how to choose the best neural network architecture. i.e., to choose the neural network with the smallest approximation error. In order to understand further, one should look into the structure of the networks that have evolved over the years and the ones that are currently in use (Fig. 5).

The perceptron network proposed by Rosenblatt [36], one of the first network models, made using the neuron model of McCulloch and Pitts [13], was used to model the visual perception phenomena. The neurons used in the perceptron have a simple summation input function and a hard-limited threshold activation function or linear threshold activation function. The input values are in general real numbers and the outputs are binary. The connection structure of this perceptron is feed forward and three-layered. The first layer is a buffer, in which the sensory data is stored. The second layer is called the 'feature layer' and the elements of the first layer are either fully or partially connected to the second layer. The neurons from the second layer are fully connected to neurons in the output layer, which are also referred to as the "perceptron layer". The weights between the buffer and the feature layer are generally fixed and due to this reason, perceptrons are sometimes called as "single layer" networks.

A perceptron learns only when it misclassifies an input vector from the training example. i.e., if the desired output is 1 and the value produced by the network is 0, then the weights of this output neuron is increased and vice versa.

Widrow and Hoff [37] proposed another formula for calculating the output error during training: $Err_j = y_j - \Sigma w_{ij} x_i$. This learning rule was used in a neural machine called ADALINE (adaptive linear neuron).

Fig. 5 A simple two-input, one output perceptron and a bias
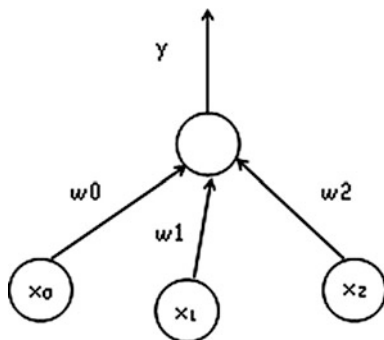
The recall procedure of the perceptron simply calculates the outputs for a given input vector using the standard summation thresholding formula and can be defined as:

$$U_j = \sum \left(X_i . w_{ij}\right) \text{for i} = (1, 2, \ldots, n) \text{for j} = (1, 2, \ldots, m)$$

where

U_j is the net input signal to each output neuron j

$X_0 = 0$ is the bias,

X = input feature vector

This perceptron can only solve problems that are linearly separable, hence could be used only to solve such examples. Still it is used due to its simplicity in structure, architecture and unconditional convergence when linearly separable classes are considered.

In order to overcome the linear separability limitations of the perceptrons, Multi-layer Perceptrons (MLPs) were introduced. A MLP consists of an input layer, at least one intermediate layer or "hidden layer" and one output layer. The individual neurons of layers are either fully or partially connected to the neurons of the next layers depending on the type and architecture of the network.

The MLPs were actually put into use only after the development of learning algorithms for multi-layer networks. i.e., back propagation algorithm [21, 22]. The neurons in the MLP have continuous valued inputs and outputs, summation input function and non-linear activation function. A MLP with one hidden layer can approximate any continuous function to any desired accuracy, subject to a sufficient number of hidden nodes. Finding the optimal number of hidden nodes for different kinds of problems has been tried out by research work and the following are considered to be the latest techniques in finding the optimal number of hidden nodes.

1 Growing neural networks: Training starts with the small number of hidden nodes and depending on the error calculated the number of the hidden nodes might increase during the training procedure.
2 The weak connections and the neurons connected by weak connections are removed from the network during the training procedure. After removing the redundant connections and nodes the whole network is trained and the remaining connections take the functions of the pruned ones. Pruning may be implemented through learning-with-forgetting methods.

Growing and pruning could be applied to input neurons hence, the whole network could be made dynamic according to the information held in the network or to be more precise according to the requirement of the nodes needed to hold the information in the data set.

Ever since the introduction of MLPs, research with diversified approaches has been conducted to find out the best network architecture, the network paradigm with the smallest approximation error for different kind of problems. Such research

conducted by different interested groups and teams has proved that certain classes of network paradigms be best used to solve particular set of problems. One such approach is the use of Self-Organizing Maps for "data mining" purposes in "knowledge discovery".

4   Knowledge Extraction from ANNs

In the past decade, another important AI topic has emerged in the form of knowledge extraction using trained neural networks. Knowledge processing is performed in a "black box" approach with the trained neural network models. Some of ANN's black box related issues are discussed in Chap. "Order in the Black Box: Consistency and Robustness of Hidden Neuron Activation of Feed Forward Neural Networks and its Use in Efficient Optimization of Network Structure". Meanwhile, the following approaches are currently used in order to extract or interpret the symbolic knowledge encoded in the structure of trained network models [8]:

(a) Decompositional: Each neuron is examined and the knowledge extracted at this level is then combined to form the knowledge base of the entire network.
(b) Pedagogical: Only the network input/output behaviour is observed, viewed as a learning task, in which the target concept is the function, computed by the network.

Today artificial neural networks have been recognized in the commercial sphere too as a powerful solution for building models of systems or subjects you are interested in, just with the data you have without knowing what's happening internally. This is not possible with the conventional computing and currently there are plenty of areas where ANN applications are commercially available i.e., Classification, Business, Engineering, Security, Medicine, Science, Modelling, Forecasting and Novelty detection to name a few.

AI not only tends to replace the human brain in some representational form, it also provides the ability to overcome the limitations faced in conventional (sequential) computing and in [10] Kasabov (1995) classifies the main paradigms adapted to achieve AI as the following:

1   Symbolic—Based on the theory of physical symbolic systems proposed by Newel and Simon [38] symbolic AI is further classified into two and they are:

 (i) A set of elements (or symbols) which can be used to construct more complicated elements or structures and
(ii) A set of processes and rules which, when applied to symbols and structures, produce new structures.
    Symbolic AI systems in the recent past have been associated with two issues, namely, representation and processing (reasoning). The currently

developed symbolic systems or models solve AI problems without following the way, how humans think, but produce similar results. They have been very effective in solving problems that can be represented exactly and precisely and have been successfully applied in natural language processing, expert systems, machine learning, and modelling cognitive processes. At the same time, symbolic AI systems have very limited power in handling inexact, uncertain, corrupted, imprecise or ambiguous information (Fig. 6).

2 Subsymbolic—Based on Neurocomputing explained by Smolenski [39] Subsymbolic, intelligent behaviour is performed at a subsymbolic level, which is higher than the neuronal level in the brain but different from the symbolic paradigms. Knowledge processing is carried out by changing states of networks constructed of small elements called neurons that are similar to the biological neurons. A neuron or a collection of neurons could be used to represent a micro feature of a concept or an object. It has been shown that it is possible to design an intelligent system that achieves the proper global behaviour even though all the components of the system are simple and operate on purely local information. ANNs, also referred to as connectionist models, are made possible by subsymbolic paradigms and have produced good results especially, in the last two decades. ANN applications i.e., pattern recognition, image and speech processing, have produced significant progress.

Increasingly, Fuzzy systems are used to handle inexact data and knowledge in expert systems. Fuzzy systems are actually rule-based expert systems based on fuzzy rules and fuzzy inference. They are powerful in using inexact, subjective, ambiguous, data and vague knowledge elements. Many automatic systems (i.e., automatic washing machines, automatic camera focusing, control of transmission are a few among the many applications) are currently in the market. Fuzzy systems can represent symbolic knowledge and also use numerical representation similar to the subsymbolic systems.

Fig. 6 Usability of different methods for knowledge engineering and problem solving depending on the availability of data and expertise on a problem based on [10] p. 67
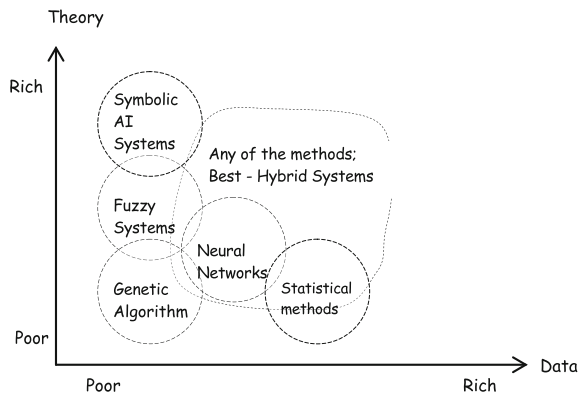
Symbolic and subsymbolic models could interact in the following ways in knowledge processing:

1 Developed and used separately and alternatively.
2 Hybrid systems with both symbolic and subsymbolic systems.
3 Subsymbolic systems could be used to model pure symbolic systems.

With that introduction to ANN initiatives, architectures, their components and hybrid systems, the remaining chapters of the book look at more recent ANN applications in a range of problem domains and are presented under three categories, namely, (1) Networks, structure optimisation and robustness (2) Advances in modelling biological and environmental systems, and (3) Advances in modelling social and economic systems.

References

1. S. Furber, Billion pound brain project under way. BBC by Fergus Walsh, 2 Oct 2013, http://www.bbc.co.uk/news/health-24428162. Accessed 18 Oct 2015
2. H. Markram, Billion pound brain project under way. BBC by Fergus Walsh, 7 Oct 2013, http://www.bbc.co.uk/news/health-24428162. Accessed 18 Oct 2015
3. B. Obama, Obama proposes brain mapping project. BBC, 2 April 2013, http://www.bbc.co.uk/news/science-environment-22007007. Accessed 18 Oct 2015
4. S.-i. Amari, Foreword, in *Foundations of Neural Networks, Fuzzy systems and Knowledge Engineering*, ed. by N.K. Kasabov (USA, Bradford, 1995), p. xi
5. J. Holland, Adaptation in Natural and Artificial Systems. The University of Michigan, 1975
6. G. Matsumato, The Brain can Acquire its Algorithm in a Self-Organized Fashion, in *Proceedings of the ICONIP/ANZIIS/ANNES'99 International Workshop*, Perth, Australia, 22–23 Nov 1999, p. 116
7. T.M. Mitchell, Machine learning and Data Mining. *Communications of the ACM,* Nov 1999, vol 42, No. 11
8. L. Fu, Knowledge discovery basedon neural networks. *Communications of the ACM, Nov 1999,* vol. 42. No. 11, pp. 47–50
9. I. Aleksander, I compute therefore I am. London BBC Science and Technology, 1999, http://news.bbc.co.uk/2/hi/science/nature/166370.stm. Accessed 27 Aug 2015
10. N.K. Kasabov, *Foundations of Neural Networks*, A Bradford Book, (The MIT Press, Fuzzy systems and Knowledge Engineering, Cambridge, Massachusetts, London, England, 1995), 581 pp
11. T. Kohonene, G. Deboeck, *Visual Explorations in Finance with Self-organizing Maps*, (Springer, London, 1998), 258 pp
12. W. Pitts, W.S. McCulloch, How we know univesrsals, The perception of auditory and visual forms. Bull. Math. Biophys. **9**(3), 127–147 (1947)
13. W.S. McCulloch, W.H. Pitts, A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. **5**, 115–133 (1943)
14. B. Widrow, Generalization and information storage in networks of Adaline 'Neurons', in *Self-Organizing Systems, symposium proceedings*, ed. by M.C. Yovitz, G.T. Jacobi, G. Goldstein (Spartan Books, Washington, DC, 1962), pp. 435–461
15. B. Widrow, Review of "Adaptive Systems of Logic Network and Binary Memories", Trans. Electron. Comput.: IEEE J. Aleksander **EC16**(5), 710–711, (1967)

16. B. Widrow, J.B. Angell, Reliable, trainable networks for computing and control. Aerosp. Eng. **21**(9), 78–123 (1962)
17. B. Widrow, M.E. Hoff, Associative Storage and Retrieval of Digital Information in Networks of Adaptive `Neurons'. Biol. Prototypes Synth. Syst. **1**, 160 (1962)
18. M.L. Minsky, S. Papert, *Perceptrons: An Introduction to Computational Geometry* (MIT Press, Cambridge, MA, 1969)
19. S. Amari, Neural theory of association and concept formation. Biol. Cybern. **26**, 175–185 (1977)
20. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. U.S.A. **79**, 2554–2558 (1982)
21. D.E. Rumelhart, G. Hinton, R. Williams, Learning internal representation by error propagation, in D.E. Rumelhart, J.L. McClelland, PDP Research Group (eds), *Parallel Distributed Processing Exploration in the Microstructure of Cognition: Foundations*, vol. 1 (MIT Press, Cambridge, MA, 1986) p. 1
22. P. Werbos, Backpropagation through time: What it does and how to do it., Proc. IEEE **87**, 10 (1990)
23. G.A. Carpenter, S. Grossberg, ART 2: Stable self-organization of pattern recognition codes for analog input patterns. Appl. Opt. **26**, 4919–4930 (1987)
24. G.A. Carpenter, S. Grossberg, A massively parallel architecture for a self-organizing neural pattern recognition machine. Comput. Vision, Graph., Image Proc. **37**, 54–115 (1987)
25. T. Kohonen, Self-organized formation of topologically correct feature maps. Biol. Cybern. **43**, 59–69 (1982)
26. B. Kosko, Bidirectional associative memories. IEEE Trans. Syst., Man Cybern. **18**, 49–60 (1988)
27. T. Moody, C. Darken, Fast learning in networks of locally tuned processing units. Neural Comput. **1**, 281–294 (1989)
28. J. Taylor, C. Mannion, *New Developments in Neural Computing* (Adam Hilger, Bristol, England 1989)
29. I. Aleksander, Neural Computing Architectures. The Design of Brain-like Machines, in I. Aleksander, H. Morton (eds.), (MIT Press, Cambridge, MA, 1990) (An Introduction to Neural Computing, London, Chapman & Hall., 1989)
30. T. Yamakawa, Pattern recognition hardware system employing a fuzzy neuron, in *Proceedings of the International Conference on Fuzzy Logic and Neural Networks,* Iizuka, Japan, July 1990, pp. 943–948
31. T. Furuhashi, T. Hasegawa, S. Horikawa et al., An adaptive fuzzy controller using fuzzy neural, in *networks. In: Proceedings of Fifth International Fuzzy Systems Association World Congress, IEEE* (1993), pp. 769–772
32. W. Freeman, C. Skarda, Spatial EEG patterns, non-linear dynamics and perception: The neo-Sherringtonian view. Brain Res. Rev. **10**(10), 147–175 (1985)
33. K. Kaneko, Clustering, coding, switching, hierarchical ordering, and control in network of chaotic elements. Physica **41D**, 137–172 (1990)
34. R. Borisyuk, A. Kirillov, Bifurcation analysis of neural network model. Biol. Cybern. **66**, 319–325 (1992)
35. B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Approach to Machine Intelligence* (Englewood Cliffs, NJ, Prentice-Hall, 1992)
36. F. Rosenblatt, The perceptron: A probabilistic model for information storage and organization in the Brain. Psychol. Rev. **65**, 386–408 (1958)
37. B. Widrow, M.E. Hoff, Adaptive switching circuits. IRE WESCON Convention Rec. N.Y. **4**, 96–104 (1960)
38. A. Newell, H.A. Simon, *Human Problem Solving, Englewood Cliffs* (Prentice Hall, NJ, 1972)
39. P. Smolenski, Tensor product variable binding and the representation of symbolic structures in connectionist systems. Artif. Intell. **46**, 159–216 (1990)