Neural Networks and Micromechanics

Bearbeitet von
Ernst Kussul, Tatiana Baidyk, Donald C Wunsch

1. Auflage 2009. Buch. x, 221 S. Hardcover
ISBN 978 3 642 02534 1
Format (B x L): 15,5 x 23,5 cm
Gewicht: 514 g

Weitere Fachgebiete > Technik > Biotechnologie > Biotechnologie: Mikrotechnologie, Nanobiotechnologie

schnell und portofrei erhältlich bei

beck-shop.de
DIE FACHBUCHHANDLUNG

Die Online-Fachbuchhandlung beck-shop.de ist spezialisiert auf Fachbücher, insbesondere Recht, Steuern und Wirtschaft. Im Sortiment finden Sie alle Medien (Bücher, Zeitschriften, CDs, eBooks, etc.) aller Verlage. Ergänzt wird das Programm durch Services wie Neuerscheinungsdienst oder Zusammenstellungen von Büchern zu Sonderpreisen. Der Shop führt mehr als 8 Millionen Produkte.
Chapter 2
Classical Neural Networks

During the last few decades, neural networks have moved from theory to offering solutions for industrial and commercial problems. Many people are interested in neural networks from many different perspectives. Engineers use them to build practical systems to solve industrial problems. For example, neural networks can be used for the control of industrial processes.

There are many publications that relate to the neural network theme. Every year, tens or even hundreds of international conferences, symposiums, congresses, and seminars take place in the world. As an introduction to this theme we can recommend the books of Robert Hecht-Nielsen [1], Teuvo Kohonen [2], and Philip Wasserman [3], and a more advanced book that is oriented on the applications of neural networks and is edited by A. Browne [4]. In this book it is assumed that the reader has some previous knowledge of neural networks and an understanding of their basic mechanisms. In this section we want to present a very short introduction to neural networks and to highlight the most important moments in neural network development.

2.1 Neural Network History

Attempts to model the human brain appeared with the creation of the first computer. Neural network paradigms were used for sensor processing, pattern recognition, data analysis, control, etc. We analyze, in short, different approaches for neural network development.

2.2 McCulloch and Pitts Neural Networks

The paper of McCulloch and Pitts [5] was the first attempt to understand the functions of the nervous system. For explanation, they used very simple types of neural networks, and they formulated the following five assumptions according to the neuron operation:
1. The activity of the neuron is an “all-or-none” process.
2. A certain fixed number of synapses must be excited within the period of latent addition in order to excite a neuron at any time, and this number is independent of previous activity and position of the neuron.
3. The only significant delay within the nervous system is synaptic delay.
4. The activity of any inhibitory synapse absolutely prevents excitation of the neuron at that time.
5. The structure of the net does not change with time.

The McCulloch-Pitts neuron is a binary device (two stable states of a neuron). Each neuron has a fixed threshold, and every neuron has excitatory and inhibitory synapses, which are inputs of the neuron. But if the inhibitory synapse is active, the neuron cannot turn on. If no inhibitory synapses are active, the neuron adds its synaptic inputs. If the sum exceeds or equals the threshold, the neuron becomes active. So the McCulloch-Pitts neuron performs simple threshold logic.

The central result of their paper is that the network of the simplest McCulloch-Pitts neurons can realize any finite complex logical expression and compute any arithmetic or logical function. It was the first connectionist model.

2.3 Hebb Theory

Hebb tried to work out the general theory of behavior [6]. The problem of understanding behavior is the problem of understanding the total action of the nervous system, and vice versa. He attempted to bridge the gap between neurophysiology and psychology. Perception, learning in perception, and assembly formation were the main themes in his scientific investigations. Experiments had shown perceptual generalization. The repeated stimulation of specific receptors will lead to the formation of an “assembly” of association-area cells which can act briefly as a closed system. The synaptic connections between neurons become well-developed. Every assembly corresponds to any image or any concept. The idea that an image is presented by not just one neuron but by an assembly is fruitful. Any concept may have different meanings. Its content may vary depending on the context. Only the central core of the concept whose activity may dominate in the system as a whole can be almost unchangeable. The possible presentation of an image or concept with one neuron deprives this concept of its features and characteristics. The presentation with a neuron assembly makes possible a concept or image description with all features and characteristics. These features can be influenced by the context of the situation where the concept is used. For example, we create the model of the concept “building” (Fig. 2.1). We can observe the building from different positions. A perceived object (building) consists of a number of perceptual elements. We can see many windows or a door.

But from different positions there are walls and a roof of this building. In an assembly that is the model of the concept “building,” a set of neurons corresponds
to the walls, other neurons correspond to windows, and others correspond to the white color of the walls, and so on. The more frequently perceived features of the building form the core of the assembly, and rare features create a fringe of the assembly (Fig. 2.2).

Due to the fringe of the assembly, different concepts may have a large number of associations with other concepts. “Fringe” systems were introduced by Hebb to explain how associations are provided. Different circumstances lead to varying fringe activity. If it is day, the white color of the building will be observed, and in the model the neuron set that corresponds to color will be excited. “Core” is the most connected part of the assembly. In our example, the core will be neurons that correspond to walls and windows. The conceptual activity that can be aroused with
limited stimulation must have its organized core, but it may also have a fringe content, or meaning, that varies with the circumstances of arousal.

An individual cell or neuron set may enter into more than one assembly at different times. The single assembly or small group of assemblies can be repeatedly aroused when some other activity intervenes. In vision, for example, the perception of vertical lines must occur thousands of times an hour; in conversation, the word “the” must be perceived and uttered with very high frequency; and so on.

2.4 Perceptrons

Perceptrons, the architecture of classical neural networks, were invented and proposed by Frank Rosenblatt in 1957 [7, 8]. The perceptron function is a classification of different patterns. A pattern can be considered as a point in \( n \)-dimensional space (where \( n \) coordinates correspond to different features of the object to be classified).

In the most common case, the perceptron was presented as a structure with one layer of neurons that are connected with inputs of the system. These connections have the weight coefficients, which can be changed during the training process. The goal is to find a set of weights \( w_0, w_1, \ldots, w_n \) such that the output of the perceptron is 1 if the input pattern vector belongs to class 1, and 0 if the pattern vector belongs to class 0. The weights are modified in accordance with the perceptron learning rule (or law). The neural network was trained using supervised training. In other words, for each input \( X \) to the network, the correct output \( Y \) also was supplied.

The perceptron was one of the first neural network paradigms, and it is still used occasionally. Their simple device structure and fast training convergence made Rosenblatt perceptrons attractive to researchers. Rosenblatt stressed that perceptrons were not developed to solve any practical task of pattern recognition or artificial intelligence. It was the model of a human brain rather than an applied technical device. However, it was clear that perceptrons could be used in practical applications, too.

Often, the Rosenblatt perceptron is considered a one-layer perceptron [9, 10]. Three-layered Rosenblatt perceptrons usually are mentioned in an historical context [11], though Rosenblatt investigated mainly three-layered perceptrons. It is interesting to build new classifiers on the base of the three-layered Rosenblatt perceptron and examine whether they can compete with the modern neural classifiers.

Analyzing the principal deficiencies of perceptrons, Rosenblatt mentioned the following [8]:

1. An excessively large system may be required.
2. The learning time may be excessive.
3. The system may be excessively dependent on external evaluation during learning.
4. The generalization ability is insufficient.
5. The ability to separate essential parts in a complex sensory field (analytic ability) is insufficient.

These points should be revised in the context of modern computer capabilities. Currently, computers cannot implement neural networks comparable with the human brain, which contains many billions of neurons, but it is possible to simulate the neuron structures containing up to, and in some cases larger than, a million neurons. In this case, it is interesting to know how the number of associative neurons influences Rosenblatt perceptron performance.

We studied and described several modifications of Rosenblatt perceptrons and experiments with them (O. Makeyev participated in this investigation) [12–19]. These experiments show that it is possible to overcome the above-mentioned problems using modern hardware. In the experiments, the number of associative neurons was changed from 1,000 to 512,000. The proposed perceptrons were tested on a benchmark MNIST data set for handwritten digits recognition [20, 21]. The performance of the modified Rosenblatt perceptron, having 512,000 neurons, is 99.2% on this database. As computer technology improves, larger capacity recognizers become feasible and higher recognition rates become possible. There are data about different classifiers’ performances on this database. The best classifier on this database shows 99.3% [21].

Bernard Widrow was working along similar lines using systems known as Adalines (ADAptive LINear Element - a single processing unit with threshold non-linearity) [22, 23]. Widrow, along with his graduate student M. Hoff, proposed the Widrow/Hoff learning law or delta rule for neural network training. In Widrow learning, the goal is to find the best possible weight vector (for a very simple type of processing element) in terms of a least mean squared error performance function criterion. This learning rule is one of the most powerful and guarantees finding this optimum weight vector from any starting point.

But in 1969, Marvin Minsky and Seymour Papert attacked neural network research. They used predicates to describe the perceptron work [24]. In particular, the following points are critical remarks concerning perceptron functioning:

1. The idea of thinking about classes of geometrical objects as classes of \( n \)-dimensional vectors \( (a_1, \ldots, a_n) \) loses the geometric individuality of the patterns and leads only to a theory that can do little more than count the number of predicates.
2. Little attention has been paid to the size, or more precisely, the information content, of the parameters \( (a_1, \ldots, a_n) \). Some examples exist where the ratio of the largest to the smallest of the coefficients is meaninglessly large. In some cases, the information capacity needed to store \( a_1, \ldots, a_n \) is even greater than that needed to store the whole class of figures defined by the pattern.
3. Closely related to the previous point is the problem of time of convergence in a learning process.
Minsky and Seymour pointed out that a single-layer network can only classify data sets that are linearly separable and hence cannot solve problems such as the EXCLUSIVE OR (XOR) (Fig. 2.3). The input vectors (0, 0) and (1, 1) belong to class \( b \), while the input vectors (0, 1) and (1, 0) belong to class \( c \). It is clear that there is no linear decision boundary that can classify all four points correctly. This problem can be generalized to \( n \) dimensions, when it is known as the \( n \)-bit parity problem.

This was the sunset of neural network investigation. Artificial intelligence methods were developed and took the place of neural network investigation. From 1969 until 1982, neural network investigations had to go underground in the US, but in the Soviet Union, Europe, and Japan these investigations continued [25–30]. For example, in 1972–1975, the first autonomous transport robot was created in the USSR [26, 27]. The robot demonstrated obstacle avoidance and purposive movement in a natural environment. TAIR was a three-wheel power barrow equipped with a system of sensors as rangefinder and tactile sensors. It was controlled by a hardware-implemented neural network (the network nodes were realized as special transistor electronic circuits; connections between nodes were realized as resistors). TAIR is presented in Fig. 2.4.

While in motion, the robot was supposed to avoid obstacles such as people, trees, park benches, and so on. Coordinates of a point in the environment gave the target of the robot’s motion. It was evident from the behavior of TAIR that, in principle, it is possible to create an entirely autonomous robot operated by a hardware-implemented neural network. At the same time, it showed the overall complexity of organization of the robot’s interaction with the natural environment, as well as the necessity of using trainable neural networks.

### 2.5 Neural Networks of the 1980s

In the early 1980s, a new wave of interest arose due to the publication of John Hopfield [31], a researcher in the field of biophysics. He described the analogy between Hebb’s neural network model and the certain class of physical systems. His efforts allowed hundreds of highly qualified scientists and engineers to join in...
the neural network investigation. At this time, the DARPA (Defense Advanced Research Projects Agency) project was initiated.

Around 1986, the new term “neurocomputer” appeared. Many international conferences on neural networks, neurocomputing, and neurocomputers took place all over the world. Hundreds of firms dedicated to neural network technology development and production were established. For example, the neurocomputer Mark III was built at TRW, Inc. during 1984–1985, followed by Mark IV [1]. In 1988, the firm HNC (Hecht-Nielson Corporation) produced the neurocomputer “ANZA plus,” which can work together with PC 386, Sun. In the same year, the neurocomputer Delta II was produced by the firm SAIC.

In the department of network system of information processing, at the Institute of Cybernetics, Kiev, Ukraine, the first neurocomputer “NIC” was created in 1988–1989 [32, 33] under the direction of Ernst Kussul. This neurocomputer is presented in Fig. 2.5. It was built on a domestic element base and was a personal computer add-on. Kussul put forward and analyzed a new neural network paradigm, which enabled the creation of neuron-like structures. These structures are known as associative-projective neuron-like networks [34–36].

After that, in 1991–1992, the Ukrainian-Japanese team created a new neurocomputer that used a more advanced element base. It was named “B-512,” and it is presented in Fig. 2.6. Kussul and his collaborators and disciples Tatiana Baidyk, Dmitrij Rachkovskij, Mikhail Kussul, and Sergei Artykutsa participated in the neurocomputer development together with the Japanese investigators from “WACOM,” Sadao Yamamoto, Masao Kumagishi, and Yuji Katsurahira.
The latest neurocomputer version was developed and tested on image recognition tasks. For example, the task of handwritten words recognition was resolved on this neurocomputer [37].

**Fig. 2.5** First neurocomputer “NIC” developed at the Institute of Cybernetics, Kiev, Ukraine

**Fig. 2.6** The neurocomputer “B-512” was created in 1992
2.6 Modern Applications of Neural Network Paradigms

There are different approaches to neural network presentation, and different paradigms of neural networks have been developed. Among them, the most popular have been the Hopfield neural network [38, 39], adaptive resonance theory developed by S. Grossberg and G. Carpenter [40, 41], Kohonen neural networks [42], Fukushima cognitron and neocognitron [43], backpropagation [44–46], and adaptive critic design [47, 48].

2.6.1 Hopfield Neural Networks

Hopfield described his neural networks in 1982 [38]. The structure of this neural network has neural processing elements. The output of every processing element is the input of other neural processing elements. The transfer function of every processing element is:

\[
X^{t+1}_i = X^t_i, \quad \text{if} \quad \sum_{j=1}^{n} w_{ij} X^t_j = T_i, \\
X^{t+1}_i = X^t_i, \quad \text{if} \quad \sum_{j=1}^{n} w_{ij} X^t_j > T_i, \\
X^{t+1}_i = X^t_i, \quad \text{if} \quad \sum_{j=1}^{n} w_{ij} X^t_j < T_i,
\]

for \( i = 1, \ldots, n \),

where \( w_{ij} \) is the weight of the input with the restrictions \( w_{ij} = w_{ji} \) and \( w_{ii} = 0 \), and \( T_i \) is the threshold. The behavior of the Hopfield network is organized in such a way as to minimize the energy function. No matter what its initial state, the Hopfield network always converges to a stable state in a finite number of processing element update steps.

There are many developments of the Hopfield network which are used in different applications. For example, the Hopfield network is a base of a hybrid Hopfield network-simulated annealing algorithm used for frequency assignment in satellite communications systems [49]. They use a fast digital Hopfield neural network, which manages the problem constraints, hybridized with simulated annealing, which improves the quality of the solutions obtained. Another example is a Hopfield neural network application for general predictive control [50]. In this case, the Hopfield neural network is used to solve quadratic optimizing problems. Existing predictive control methods are very complex and time consuming. With this proposition of a Hopfield neural network application and the development of a neural network chip, the method has a promising future in industry. The Hopfield neural network is used in information retrieval systems, too [51]. In recent years,
with the rapid development of the Internet and easy access to a large amount of information on it, information retrieval has become more indispensable for picking out useful data from the massive resources. With the heuristic function of the Hopfield network, this model is used in query expansion, and therefore can solve the problems of information overload and word mismatch to some extent.

2.6.2 Adaptive Resonance Theory (ART)

Stephen Grossberg and Gail Carpenter [40, 41, 52] developed and introduced a variety of neural network paradigms. The most popular is adaptive resonance theory, which finds an application for different task solutions.

For example, driving safety is a very important consideration for the automotive industry and for consumers. The methods for improving driving safety can be roughly categorized into passive or active. Passive means (e.g., seatbelts, airbags, etc.), which have significantly reduced traffic fatalities, were originally introduced to diminish the degree of injury during an accident. Active means, on the other hand, are designed to prevent accidents in the first place. A driver assistance system is one kind of active system that is intended to alert a driver to the potential of a dangerous situation as soon as possible. Detecting critical changes in the driving environment is an important task in driver assistance systems. A computational model was developed for this purpose [53]. This model includes three components: sensory, perceptual, and conceptual analyzers. They use visual sensors (cameras and video camcoders). Each video sequence was down-sampled to a frame rate of 5 Hz to reduce the processing load on the computer. This frame rate is also fast enough for a driver to respond to any of the environmental changes. The size of each input image \((320 \times 240 \text{ pixels})\) was reduced to \(160 \times 120 \text{ pixels}\) to reduce the processing time. A number of video sequences were collected and categorized into seven classes, referred to as the right-lane-change, left-lane-change, tunnel entry, tunnel exit, freeway entry, freeway exit, and overpass ahead. Each class was further divided into two groups, termed the “day” and “night” groups. A vision system for detecting critical changes in driving was developed. In this framework, both temporal and spatial information are extracted from input video sequences. The extracted information serves as input stimuli to a spatiotemporal attention neural network. The attention pattern associated with the focus, together with the location and direction of motion of the pattern, form what Chiung-Yao Fang et al. call a categorical feature. Thereafter, based on this feature, the class of the attention pattern and, in turn, the change in driving environment corresponding to the class is determined using a configurable adaptive resonance theory (CART) neural network. This is the work of the conceptual analyzer. Various changes in the driving environment, both in the daytime and at night, have been tested [53].

Adaptive resonance theory was applied to resolve large-scale traveling salesman problems (TSP) [54]. TSP is a very difficult optimization issue in the field of operations research. Using adaptive resonance theory and local optimization to
divide and conquer large scale TSP has the advantages of increased scalability and parallelism.

An ART-2 network was used to develop a strategy for the navigation of mobile robots in uncertain indoor environments [55]. More exactly, a modified ART-2 network was put forward to identify the surrounding environment correctly for mobile robots. Path planning is one of the vital tasks in the navigation of autonomous mobile robots and may be divided into two categories: global path planning based on a priori complete information about the environment and local path planning based on sensor information in uncertain environments where the size, shape, and location of obstacles are unknown. Local path planning could be called reactive strategies. The neural networks and fuzzy controllers have proved to perform well in reactive navigation applications. Computer simulations were made for design strategies of environment classifiers based on a modified ART-2 network and fuzzy controller. One more example of an ART application is connected with a mobile vehicle [56]. In this case, an ART network is used for image processing. On the robot’s path are several labels, the letters L, R, B, F, and S, which represent turning left or right, moving backward or forward, and stopping. ART is used to recognize them and give out the signal to control a mobile vehicle. Other experiments were conducted involving obstacle avoidance. Eight obstacle patterns were selected to train the ART network. The potential to add new categories is very important for this type of task.

In another study [57], the ART family of neural networks was used to develop a speaker recognition system. This system consists of two modules: a wavelet-based feature extractor and a neural-network-based classifier. Performance of the system has been evaluated using the gender recognition and speaker recognition problems. In the gender recognition problem, the highest accuracy was 90.33%; in the speaker recognition problem, ART-based classifiers have demonstrated recognition accuracy of 81.4%.

### 2.6.3 Self-Organizing Feature Map (SOFM) Neural Networks

Teuvo Kohonen published his first articles in the seventies [29]. He applied a specific type of neural network – Self-Organizing Feature Map (SOFM) designs with the following self-organization training principles. The principal idea is that a set of processing elements arrange their weights in such a way that they are distributed in space with a density approximately proportional to the probability density of the input vector.

This approach found a place in modern investigations and applications. For example, a Kohonen self-organizing map can be used for unsupervised segmentation of single-channel magnetic resonance (MR) images [58]. The introduction of advanced medical techniques, such as MRI, has dramatically improved the quality of the diagnosis and treatment of brain pathologies. The image information in such systems is complex and has a great number of dimensions. The availability of
segmentation techniques is useful in assisting medical experts in the diagnosis and treatment of tumors. The segmentation of single-channel magnetic resonance images is a daunting task due to the relatively small amount of information available at each pixel site. This method has been validated on both simulated and real images of volunteers and brain tumor patients. This is the first step in developing a fully automatic segmentation method.

The other area where Kohonen maps found an application is for a distributed measurement system for water quality monitoring [59]. Water quality monitoring of rivers and seas represents an important task of life-quality assessment. This monitoring is characterized by multi-parameter measurement capabilities. The main parameters associated with water quality inspection can be classified in three categories: physical, chemical, and biological parameters. They use an advanced processing of data sensors based on auto-associative neural networks (Kohonen maps) in order to offer a global water quality representation for a large monitored area.

One more interesting application of Kohonen maps is feature selection for object extraction [60]. Selecting a set of features that are optimal for a given task is a problem that plays an important role in pattern recognition, image understanding, and machine learning. Li Pan et al. used Kohonen maps for continuous data discretization in texture-recognition tasks. As the test task, they used tree recognition from aerial images.

Also, Kohonen maps were used for color image compression [61]. With the development of multimedia technology and the Internet, image communication including transmission, display, and storage at high speed has become increasingly important. In this case, the hardware design for a neural-network-based color image compression was developed. Compression using neural networks is advantageous due to their features such as inherent parallelism, regular topology, and their relatively small number of well-defined arithmetic operations involved in their learning algorithms. So, VLSI implementation of Kohonen’s map neural network is well suited to color image compression due to its topological clustering property. In this case, similar colors are clustered together and can be represented by one color. The time complexity of the proposed scheme is linear in the image size. With ASIC implementation the compression time is only a few milliseconds for images of sizes up to $512 \times 512$ pixels.

Extended Kohonen networks were used for the pose control of microrobots in a nanohandling station [62]. These are only a few examples among many others.

### 2.6.4 Cognitron and Neocognitron

Kunihico Fukushima developed and proposed the *cognitron* neural network [30], which is an example of a hierarchical network [43]. It was initially proposed as a neural network model of the visual system that has a hierarchical multilayered architecture similar to the classical hypothesis of Hubel and Wiesel [63, 64]. The model consists of S- and C-cells. S-cells resemble simple cells of the primary
visual cortex, and C-cells resemble complex cells. The layers of S- and C-cells are arranged alternately in a hierarchical structure. S-cells feature extracting cells and have variable input connections, which can be modifiable during training. C-cells have fixed, non-modifiable input connections.

In recent years, these ideas have found new applications [65–67], and [68]. In Japan, at Toyota Central R&D Labs., Inc., the neocognitron is used for position detection and vehicle recognition [65]. This system is tolerant to deformations and shifts in the position of a vehicle. K. Fukushima [66] applies the neocognitron for handwritten digit recognition, but several new ideas have been introduced, such as the inhibitory surround in the connections from S-cells to C-cells. Fukushima also applies the neocognitron to the recognition of patterns that are partly occluded [68]. Other authors [67] have modified the neocognitron and used it for breast cancer detection. The so-called Shape Cognitron (S-Cognitron) is composed of two modules and was introduced to classify clustered micro calcifications, which generally present an early sign of breast cancer. The first module serves as a shape orientation layer and converts first-order shape orientations into numeric values. The second module is made up of a feature formation layer followed by a probabilistic neural-network-based classification layer. This system was tested on the 40-mammogram database provided by the Department of Radiology at the University of Hospital Nijmegen in the Netherlands and showed promising results.

2.6.5 Backpropagation

The backpropagation method was introduced by Paul Werbos in 1974 [46]. In 1985–1986, D. Rumelhart and others worked out and applied this mechanism for neural network training [44, 45].

At present, the backpropagation method is used to resolve different practical tasks. For example, it is used for retrieval algorithms for geophysical parameters in the retrieval of atmospheric water vapor and cloud liquid water content over oceans from brightness temperatures measured by the multi-frequency scanning microwave radiometer launched onboard satellite [69]. These studies have demonstrated the great potential of neural networks in a large variety of remote sensing and meteorological applications. For this concrete task, the multilayer perceptron (MLP) with backpropagation training algorithm was used. MLP has the ability to detect multiple nonlinear correlations from the training database. MLP has advantages over statistical regression methods [70].

The backpropagation method is used very often in medicine, for example, for classifying balance disorders using simulated sensory deficiency [71]. This task is important for medical rehabilitation of patients with sensory deficiency. Another example is connected with the classification of neck movement patterns related to Whiplash-associated disorders (WADs) using a resilient backpropagation neural network [72]. WADs are a common diagnosis after neck trauma, typically caused by rear-end car accidents. Neck movement was used as input. Rotation angle and
velocity were calculated. A principal component analysis was performed in order to reduce data and improve the backpropagation neural network performance. This method showed a predictivity of 0.89, which is a very promising result. The neck movement analysis combined with a neural network could build the basis of a decision support system for classifying suspected WADs.

Another interesting use is detecting sky in photographic images [73], which was done by collaborators of the Imaging Science Technology Laboratory, Kodak Research Laboratories, Eastman Kodak Company, USA. In their system, color classification is performed by a multilayer backpropagation neural network trained in a bootstrapping fashion to generate a belief map of sky color. Next, the region extraction algorithm automatically determines an appropriate threshold for the sky color belief map and extracts connected components. Finally, the sky signature validation algorithm determines the orientation of a candidate sky region using a physics-motivated model. With approximately half of the images containing blue sky regions, the detection rate is 96%. A feedforward neural network was structured with two hidden layers containing three and two neurons, and a single output neuron. The output of the network is a belief value between 0 and 1 for each pixel, 1 indicating a pixel highly likely to be blue sky.

In industry, the backpropagation neural network is used, for example, for fault diagnosis of circuit breakers [74]. The maintenance costs for aging circuit breakers (CBs) are significant. To reduce the cost, a condition-based maintenance is proposed. Many testing methods for characterizing the condition of a CB have been studied, such as contact travel time measurement and vibration analysis. In [74], wavelet packets are used to convert measured vibration data from healthy and defective CBs into wavelet features. Selected features highlighting the differences between healthy and faulty conditions are processed by a backpropagation neural network for classification. Testing has been done for three 66-kV CBs with simulated faults. Detection accuracy is shown to be far better than in other classical techniques such as the windowed Fourier transform, standalone artificial neural networks, or expert systems. The accuracy of detection for some defaults can be as high as 100%.

2.6.6 Adaptive Critic Design

One of the most interesting approaches to using adaptive networks to solve common problems in adaptive control and system identification is adaptive critic design [47, 48, 75]. In the broadest sense, this method is developed to handle large, noisy, nonlinear problems in real time. Many applications of neural networks to control were limited to the use of static or feedforward networks, adapted in relatively simple ways [75]. Many applications of that sort encountered problems such as limited application or slow adaptation.

Adaptive critic designs may be defined as designs that attempt to approximate dynamic programming in the general case. Dynamic programming is the method for finding an optimal strategy of action over time in a noisy, nonlinear
The user supplies a utility function $U$ and a stochastic model $F$ of the environment to be controlled. Dynamic programming is used to solve for another function, $J$, which serves as a secondary or strategic utility function. The key theorem is that any strategy of action that maximizes $J$ in the short term will also maximize the sum of $U$ over all future times. Adaptive critic designs are defined more precisely as designs that include a critic network—a network whose output is an approximation to the $J$ function, or to its derivatives, or to something very closely related to the two. There are different realizations of this approach; it is possible to say that an adaptive critic family of methods has developed.

For example, adaptive-critic-based neural networks have been used to design a controller for a benchmark problem in aircraft autolanding [76], to steer an agile missile [77], to build neurocontrollers for turbo generators in a multimachine power system [78], and to construct new learning methods (creative learning) for intelligent autonomous mobile robots [79].

We plan to use adaptive critic design in our future investigations of micromechanics and microfabrics control.

### 2.7 RTC, RSC, LIRA, and PCNC Neural Classifiers

We have worked out effective neural network classification systems. They have been developed since 1970 and used as control systems for mobile autonomous robots, in texture recognition tasks, in voice-based identity verification tasks, in handwriting and face recognition, and in the new micromechanics area. The most interesting neural classifiers are Random Threshold Classifier (RTC) [80, 81], Random Subspace Classifier (RSC) [12, 82], Neural Classifier LIRA (Limited Receptive Area) [12–19], and PCNC Neural Classifier [83, 84]. In this book, we describe all these models, obtain results, and summarize all of the advantages of our approach.

### References

References


