

Variations of the two-spiral task

STEPHAN K. CHALUP* and LUKASZ WIKLENDT

School of Electrical Engineering and Computer Science,
The University of Newcastle, Callaghan, 2308, Australia

(Received 7 December 2005; revised 7 February 2007; in final form 29 March 2007)

The two-spiral task is a well-known benchmark for binary classification. The data consist of points on two intertwined spirals which cannot be linearly separated. This article reviews how this task and some of its variations have significantly inspired the development of several important methods in the history of artificial neural networks. The two-spiral task became popular for several different reasons: (1) it was regarded as extremely challenging; (2) it belonged to a suite of standard benchmark tasks; and (3) it had visual appeal and was convenient to use in pilot studies. The article also presents an example which demonstrates how small variations of the two-spiral task such as relative rotations of the two spirals can lead to qualitatively different generalisation results.

Keywords: Benchmarking; Classification; Generalisation; Neural networks; Support vector machines

1. Introduction

Research and development in many fields is driven by challenging tasks and applications. In machine learning, these challenges are often represented by selected data sets which are either empirically collected or artificially generated. Some of them have repeatedly been investigated and become standard benchmark tasks (Prechelt 1994; White *et al.* 1995; Blake and Merz 1998). Among them is the two-spiral classification task.

The first artificial neural networks were based on the perceptron (McCulloch and Pitts 1943, Rosenblatt 1958). Geometrically, the perceptron can be interpreted as a separating hyperplane and can be used for binary classification of linearly separable data. The simple XOR task where four points need to be classified into two classes of two points that cannot be linearly separated could not be achieved by a single layer perceptron (Minsky and Papert 1969). The XOR task triggered the next major step in the history of artificial neural networks and drove the development of multilayer perceptrons (MLPs) in numerous studies.

Through this endeavour the two-spiral task became famous because it was regarded as an extremely difficult non-linear classification task for MLPs to solve. The two-spiral task is related to Minsky and Papert's spiral problem in which the learner has to decide whether the two spirals are connected to each other or separate (Minsky and Papert 1969, 1988). The latter

*Corresponding author. Email: stephan.chalup@newcastle.edu.au

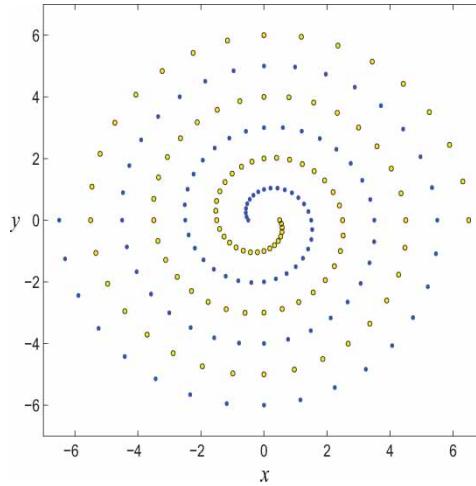


Figure 1. Two-spiral data set with 194 points.

problem has been the topic of investigation in psychology and visual perception (Howard 1974; Grossberg and Wyse 1991; Chen and Wang 2001), and the spirals as well as related radial ray patterns are similar to Klüver's hallucination form constants (Bressloff *et al.* 2001; Bressloff and Cowan 2003). Spirals occur naturally in galaxies, cyclones, eddy currents in liquid, shells, DNA and many other instances.

The two-spiral classification task for artificial neural networks was proposed in the late 1980s. It was initially posted to the connectionist mailing list by Alexis P. Wieland and then included into the CMU neural network benchmark repository (White *et al.* 1995). The original two-spiral data set (Lang and Witbrock 1988) consisted of 194 data points distributed equally over two inter-locked spirals, that is, two equal spirals which are rotated π radians relative to each other. The data points were presented in Cartesian (x, y) -coordinates and plotted in a configuration where the spiral points are aligned in radial rays as in figure 1, where one spiral is represented by 97 dark dots and the other by 97 bright dots or circles. The task addresses supervised learning of binary classifiers, which use binary labelled targets to learn to which of the two spirals each input data point belongs. Generalisation ability of the trained classifier is typically tested by evaluating it on a subset of all the other pixels of the image which were not included as spiral points in the training set. The generalisation results for the two-spiral task can then nicely be visualised in images showing black and white regions as labelled by the classifier. Ideally, the resulting image is divided into two equally thickened spiral shaped regions, each of which has one of the two spirals formed by the training data as core. In section 4, a state-of-the-art learning method, namely least squares support vector machines (LSSVM) (Suykens and Vandewalle 1999a,b), is applied to show solutions to the two-spiral task and some of its variations.

2. The two-spiral data

Spiral curves have been the object of mathematical investigation over several hundreds of years. Many different types of spiral exist and can be described formally in different ways. A

general equation to describe a spiral of points $(x, y) \in \mathbb{R}^2$ in the Euclidean plane is

$$\begin{pmatrix} x \\ y \end{pmatrix} = r(\phi) \cdot \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix}, \quad (1)$$

where $r(\phi) \in \mathbb{R}^+$ is the radius and $\phi \in \mathbb{R}$ is the angle between the x axis and the radial ray connecting the origin with point (x, y) . If the expansion of the radius

$$r(\phi) = a\phi, \quad (2)$$

where $a \in \mathbb{R}^+$ is a scaling constant, is linear it is a version of Archimedes' spiral and if the radius depends exponentially on ϕ , for example,

$$r(\phi) = ae^{b\phi} \quad (3)$$

with constants $a, b \in \mathbb{R}^+$ it is called a logarithmic spiral. The latter is also known as growth spiral, equiangular spiral, spira mirabilis or Bernoulli spiral (Archibald 1918; Hilton *et al.* 1997; Weisstein 2003).

The original two-spiral data can be generated using a C-routine, which was included in the CMU repository (White *et al.* 1995; Lang and Witbrock 1988). Two parameters can be set: the density parameter controls the number of data points generated and the radius determines the range of the values generated. With the default parameters density 1 and radius 6.5, the two-spiral data set consists of 194 training points. These are divided into two classes with 97 data points, where each class corresponds to one spiral of three turns with 32 points per turn plus one endpoint. Both spirals have the same origin and orientation but are rotated relative to each other by 180° (see figure 1). The first spiral can be generated using the following equations for $i = 0, 1, \dots, 96$:

$$\begin{aligned} \phi &= \frac{i}{16} \cdot \pi \\ r &= \frac{6.5 \cdot (104 - i)}{104} \\ x &= r \cos \phi \\ y &= r \sin \phi \end{aligned} \quad (4)$$

The second spiral can be obtained by plotting $(-x, -y)$.

Several variations of the original two-spiral data are possible, and some of them have been employed as training data for classification studies in machine learning (see section 3.2). The most prominent variations include an increase in the number of training points, the addition of some randomness or an increase in the number of spirals.

3. Survey

The following sections provide a survey of some of the main studies and developments which were motivated by or involved the two-spiral data. Section 3.1 addresses studies which relied on the original two-spiral data set to develop or test a new machine-learning method. Section 3.2 includes studies which involved variations of the two-spiral data, such as spirals of higher or lower density, or noisy spirals.

3.1 Studies using the original two-spiral task

After the challenges of the two-spiral data were explored in some early studies (section 3.1.1), they triggered the development of a series of constructive methods (section 3.1.2) along with several alternative approaches and specifically adapted methods (section 3.1.3). Finally, the approaches contained in section 3.1.4 seem to represent the current state-of-the-art in solving the two-spiral task.

3.1.1 Early studies. In the late 1980s and early 1990s the two-spiral classification task was regarded as an extremely challenging non-linear classification task and was approached in several studies using different variations of neural network learning. Lang and Witbrock (1988) reported that none of their initial attempts to solve the task using a standard MLP with back-propagation training were successful. However, they succeeded in training 2-5-5-5-1 MLPs which had additional shortcut connections to every succeeding layer. Training took about 10,000–20,000 epochs. Using the quick-propagation algorithm training could be accelerated and required only 8000 epochs on average (Fahlman, 1988a,b). They found that the trained networks did not separate the (x, y) -plane in two perfect spiral regions but fitted all the training points. They argued that with the original data the task was under-constrained and demonstrated that their networks could much better approximate the spiral decision boundary if the density of points in the training set was increased by a factor of four (see section 3.2). Baum and Lang (1991) demonstrated how to train a 2-50-1 MLP by query learning. This involved a two-spiral oracle, that is, they assumed a perfect solution of the task in advance. Osowski *et al.* (1996) demonstrated that the training of a standard 2-50-1 MLP could be achieved, without preempting a solution. They employed a variation of second-order Newton optimisation and learned the two-spiral task in about 650 epochs.

3.1.2 Constructive methods. The intuitive idea that among successfully trained neural networks smaller networks might generalise better than larger networks (cf., Bartlett 1998; Anthony and Bartlett 1999) led to the development of several constructive neural network algorithms which can grow the network size. The two-spiral task became one of the favourite testbeds during this development.

When Fahlman and Lebiere proposed their cascade correlation learning architecture (CasCor), which incrementally grew a neural network unit-by-unit during training, they performed extensive tests on the two-spiral data (Fahlman and Lebiere 1990, 1991). On average, CasCor required 1700 epochs and recruited 12–19 hidden units to solve the task; that is, it was about five times faster than their previous training with the quick-propagation algorithm (Fahlman 1988a,b). The generalisation results of CasCor were slightly smoother than those obtained by Lang and Witbrock (1988) but still contained clear irregularities and jagged edges (Fahlman and Lebiere 1990, 1991).

Fahlman and Lebiere's evaluation of CasCor on the two-spiral data inspired a long sequence of follow-up studies by other researchers. Among them were the studies by Shultz and Elman (1994) who analysed CasCor nets using principal component analysis (PCA) of contributions, and by Shultz *et al.* (1995) who employed PCA of the variance-covariance matrix of contributions. Further explorations included methods to reduce the number of hidden units (Teng and Wah 1996), and some alternative approaches based on projection pursuit learning by Hwang *et al.* (1996). Advancements such as smaller solution networks and improved generalisation ability on the spirals data were reported by several authors (Treadgold and Gedeon 1997a,b, 1999), who proposed a number of variations of CasCor. Chen *et al.* (1997) proposed an algorithm which partitions the input space recursively by growing trees of MLPs

trained by backpropagation. They reported their method could solve the two-spiral task in about half the time of CasCor. Potter and De Jong (2000) compared CasCor with their method of cooperative coevolution networks which required fewer hidden units (on average 13.7) than CasCor. Thangavel and Kathirvalavakumar (2003) proposed to combine cascade learning with a weight perturbation method to construct a single hidden layer neural net. They experimentally compared their approach with the modified backpropagation method using weight extrapolation by Kamarthi and Pittner (1999). Both studies employed as one of their main test problems a simplified version of the two-spiral task, where each spiral had only 1.5 turns and 41 points. Kamarthi and Pittner (1999) were able to train a 2-8-2-1 MLP in 3798 epochs, whereas the constructive network of Thangavel and Kathirvalavakumar (2003) required 11 units, 8891 epochs, but much less time per epoch. Several other researchers developed the idea of CasCor and constructive neural network algorithms further and used the two-spiral data for evaluation (Setiono 1995; Su and Guan 2000; Lahnajärvi *et al.* 2002).

Another branch of follow-up studies grew out of a paper by Fritzke (1994), in which he proposed a version of his growing cell structures for supervised learning (SGCS) and demonstrated that they could learn the two-spiral task in about 180 epochs. An evaluation of generalisation ability of the solution networks revealed decision regions which had much smoother boundaries than the results of CasCor by Fahlman and Lebiere (1990, 1991). Bruske and Sommer (1995), Cheng and Zell (2000) and Cheng *et al.* (2006a, b) extended and advanced Fritzke's approach. Two other studies claimed they can solve the two-spiral task much faster than Fritzke's SGCS, namely Patanè and Russo (2002) reported that their fully automatic clustering system (FACS) could solve the task in about 15 epochs and Baraldi and Alpaydin (2002) claimed their constructive fully self-organising simplified adaptive resonance theory clustering networks could solve it in only two epochs.

In addition to the studies of Fahlman and Lebiere (1990, 1991), Fritzke (1994), and their follow-up studies, several alternative constructive methods were developed and tested on the two-spiral data. Among them were the networks of localised linear discriminants (Glassman, 1992) for which the two-spiral task was reported to be 'straightforward', and the constructive algorithm for real-valued examples (CARVE) by Young and Downs (1998), which was tested using linear as well as higher-order threshold activation functions. The latter version was capable of solving the two-spiral task using a single hidden-layer with eight units. This was compared with the constructive method of Holt (1994a,b) which employed between 8 and 10 hidden units with hyperspherical decision boundaries to solve the two-spiral task. Another example of a self-growing neural network architecture was developed by Fu *et al.* (2001) using extensive tests on the two-spiral task. Different versions of their self-growing modular perceptron networks were combined with learning schemes that divide the training data (Chiang and Fu 1994) and estimate the initial weights (Lee and Fu 1999).

3.1.3 Alternative approaches, specifically adapted methods and associated findings.

Koza (1992) solved the two-spiral problem with genetic programming and a population size of 10 000. Using the same genetic programming primitives, Juille and Pollack (1996) investigated the effects of co-evolution in a population of size 4096. The two-spiral task was their central benchmark. Yang and Kao (2001) proposed a new evolutionary approach and trained a 2-5-5-5-1 MLP on the two-spiral data in 27,318 generations. They noted that their approach took longer than the initial results obtained by Lang and Witbrock (1988) using backpropagation. However, Yang and Kao (2001) also suggested their method would be more stable than the genetic programming approach of Juille and Pollack (1996). Another study which followed Koza (1992) was the investigation into complementary phenotype selection for crossover by

Dolin *et al.* (2002) who approached a simplified version of the two-spiral task, which included only the 60 innermost points of the 194 original data points.

Alternative input encodings were used to facilitate training in some studies, including Alvarez-Sanchez (1999). Alvarez-Sanchez circumvented the difficulty of the two-spiral data by transforming it from Cartesian into polar coordinates. This resulted in the two spirals being represented by a set of parallel lines. Jia and Chua (1995) experimented with different input data encodings such as binary, gray and temperature coding. Chua *et al.* (1995) found advantages using the latter. No difference between binary and gray encoding was observed by Fahner and Eckmiller (1994). Zhang and Zhang (1999) experimented with a neural model which uses projections of the input vectors onto a sphere. They evaluated their approach primarily by using the two-spiral data and visual inspection of its input coverings. They claimed that upon visual inspection their results appear to surpass those of Lang and Witbrock (1988). There are several other studies which used the two-spiral data and either modified the input encoding (Singh 2001) or used suitable initialisation (Baum and Lang 1991, Denooux and Lengelle 1993, Karayiannis 1996, Lehtokangas and Saarinen 1998b).

Singh (1998b) claimed that the generalisation error of 2- h -1 MLPs trained on the two-spiral task depended on the number h of hidden units. In his experiments, the generalisation error decreased with an increase in the number of hidden units up to $h = 15$. More hidden units did not lead to further advancement. Shang and Wah (1996) evaluated their NOVEL method and different numbers of hidden units when training feed-forward networks with shortcuts. They found their method performed well with about six hidden units, and better than several other methods including CasCor. A further improvement in terms of training speed, success ratio and smaller number of weights was announced by Wah and Qian (2000), although the images visualising the generalisation ability of their resulting networks are still very different from the ideally expected spiral shaped regions. Yu *et al.* (2002) were able to train 2-12-1 MLPs using backpropagation successfully if the activation functions were made adaptive. They also used a variation of the data with 200 points per spiral. Another version of adaptive activation functions was investigated by Solazzi and Uncini (2000). Ting and Sugai (1999) compared neural networks with wavelet activation functions to other approaches on the two-spiral task. They reported that it was ‘... impossible to obtain exact training time for each method...’ but their networks appeared to train relatively quickly when compared with the other methods.

A centroid-based MLP was proposed by Lehtokangas and Saarinen (1998a), which performed better than a 2-40-1 MLP. Equipped with additional initialisation schemes, further improvement could be achieved (Lehtokangas 2000). Mizutani and Dreyfus (2002) used carefully selected learning rates together with a modified objective function and showed that five of 100 training trials with 2-30-1 MLPs could solve the two-spiral task within 150,000 epochs with 100% accuracy. Ridella *et al.* (1997) obtained good results using radial basis function (RBF) networks and their circular backpropagation networks. Later Ridella *et al.* (1999) demonstrated advantages of using circular over sigmoidal units when using their conditional class entropy networks on the two-spiral task. Kaminski and Strumillo (1997) investigated radial basis function (RBF) training, where the centres of the basis functions were selected randomly from a two-dimensional Gaussian distribution centered at the spirals’ center lines. Kuncheva (1997) employed a genetic algorithm to select the initial seed point for a RBF classifier. Chaiyaratana and Zalzal (1998) demonstrated that their hybrid RBF-MLP network requires fewer epochs than a standard RBF network when trained on the two-spiral task. They further experimented with combined genetic/supervised/unsupervised training and gave benchmark results to compare with the performance achieved in previous studies. Combinations of distance metric-based clustering and support vector classification were evaluated by Ribeiro (2002) on the two-spiral data. Huang and Zhao (2005) demonstrated that

their approach using radial basis probabilistic neural networks achieves better generalisation on the two-spiral data than standard RBFs.

Carpenter *et al.* (1992) proposed a system combining fuzzy logic and adaptive resonance theory (Fuzzy ARTMAP), which found a solution to the two-spiral task in only five epochs. A noise tolerant version of Fuzzy ARTMAP (Carpenter *et al.* 1995) was later evaluated on a noisy two-spiral task (cf., section 3.2). Xiong *et al.* (2001, 2003) tested a combination consisting of a three layered conventional neural network with 100 hidden nodes and a Gaussian fuzzy inference network. With this combined method, they obtained faster learning and better generalisation than with the standard neural network alone. A combination of fuzzy modeling and support vector machines was proposed by Chiang and Hao (2004). They demonstrated the superior generalisation ability of their system in a comparison with a 2-170-1 MLP trained by backpropagation in 100,000 epochs, a 170 unit RBF network, and Fuzzy ARTMAP. Fuzzy methods were addressed by several other studies and tested on the two-spiral task (Sun and Jang 1993, Singh 1998a, Chen and Chang 2000, deSouza *et al.* 2002).

The concept of data set partitioning or divide and conquer was employed by a large number of neural network studies which included tests on the spirals data (Whitley and Karunanithi 1991, Smieja 1996, Chen *et al.* 1997, Lu and Ito 1999, Fu *et al.* 2001, Castro *et al.* 2000, Chen and Chang 2000, de Souza *et al.* 2002). The concept of partitioning can successfully be employed to solve the two-spiral task without using neural networks at all. Bradley and Lovell (1994) proposed a rule-based inductive learning algorithm which successively splits the feature space in half. Their algorithm showed better performance than CasCor or C4.5 (Quinlan 1993). Lovell and Bradley (1996) evaluated the accuracy of their algorithm on the training set and on a test set, where the test set consisted of additional 194 points. These were located on the same spirals but between the original training points. The performance of the algorithm was further improved to 100% accuracy on both the training and the test set by adding uniform random noise to the training data (cf., Blahut 1990).

3.1.4 Recent and state-of-the-art results. Although many of the above-mentioned approaches already reported an excellent performance, there are a number of selected studies which claimed to have found methods to solve the two-spiral task efficiently and/or almost perfectly.

Sopena *et al.* (1999) trained 2-16-1 MLPs with sine as activation function in the hidden layer and obtained about 98% correct generalisation (better than CasCor) after only a short period of training.

Cioccoiu *et al.* (2001, 2002a,b) presented several different successful approaches of using RBF networks in combination with other techniques such as Kalman filters, PCA and standard MLPs. He claims his hybrid networks require fewer parameters for achieving 100% correct classification on the training set than several of the previous approaches (Cioccoiu 2002a).

Other very successful studies employed different types of support vector machines and experimented with different kernels (Ayat *et al.* 2001). Suykens and Vandewalle (1999a,b) and Suykens *et al.* (2002), for example, showed that LSSVMs with RBF kernels could solve the two-spiral task with very good generalisation and low computational cost. Compared with standard SVMs with tanh kernel, the approach of Suykens and Vandewalle (1999b) required fewer support vectors and generalised better. Suykens and Vandewalle (1999a) further reported that their method's performance on the two-spiral task was relatively robust against changes of the learning and kernel parameters and that RBF kernels should be preferred (cf., Ridella *et al.* 1999). Galleske and Castellanos (2002) proposed a method of kernel optimisation for probabilistic neural networks and claimed their approach could outperform SVMs on some variation of the two-spiral task.

3.2 Variations of the data

Most of the above-mentioned studies trained on the original data set with 97 points per spiral. However, Lang and Witbrock (1988) not only used the original data but also a version of the data with four times as many points, that is, 388 points per spiral. Training on the denser spirals took about three to four times as long. An analysis of the different stages of training and the contribution of individual network units and layers to the training outcome showed that the network could approximate the spiral shaped decision boundaries better at the centre of the spirals where the point density of the original training data is highest (cf., figure 1). Lang and Witbrock (1988) argued the original task would be under-constrained and demonstrated that their networks, when trained on the denser data set, could approximate the spiral shaped decision boundaries much better than in the original case. A small version of the spirals data with 50 points per spiral was used by Han and Moraga (1995) to demonstrate fast training ability of their ‘hybrid networks’ with different activation functions in different layers. Suykens and Vandewalle (1999b) illustrated the generalisation performance and the low number of support vectors required by their method using spirals with only 30 points. In contrast, Yu *et al.* (2002) increased the point density to 200 points per spiral and found that their MLPs with adaptive activation function generalised very well. As already demonstrated by Lang and Witbrock (1988), it should be taken into account that increasing the point density can increase the training time but it can also help to obtain better generalisation performance.

Tomandl and Schober (2001) demonstrated the good generalisation performance of their ‘Modified General Regression Neural Network’ on two variations of the spirals data: on the two-spiral data with 50 points per spiral as well as on a variation using four spirals with 100 points per spiral and three revolutions. A sequential four-spirals variant which consisted of four nested two-dimensional spirals that unfolded as temporal input sequences was used by Weir and Chen (1991).

Whitley and Karunanithi (1991) experimented with variations of the training data by incorporating not only points on the central spirals but also points located at boundaries of the decision surfaces between the spirals. Lu and Ito (1999) selected random subsets of the spiral data to train a modular neural network. Fan and Zhang (2000) used spirals with 300 points each and added some noise to the spirals’ point coordinates. They demonstrated how their proposed space-partitioning multi-layer perceptron can insert decision boundaries between the two spirals. In some of their experiments with support vector machine kernels, Ayat *et al.* (2001) used Gaussian random noise to generate the spiral point coordinates. The random noise and the training data in the latter two studies was not clearly specified. A two-spiral data set with Gaussian noise was also employed by a recent study of Du and Chen (2007) to demonstrate satisfying performance of their rectified nearest feature line segment method for pattern classification. Another version of a noisy two-spiral task with 20 sample points drawn from Gaussian clusters (standard deviation $\sigma = 0.01$) centred at each of the original 97 spiral points (*i.e.* 1940 patterns for each class) was proposed by Carpenter *et al.* (1995) and further developed by Williamson (1996) and Lee *et al.* (2004). The set-up of the noisy spirals task by Lee *et al.* (2004) was adopted from Williamson (1996), who used training data drawn from 194 anisotropic Gaussian distributions ($\sigma = 0.025$) centred along the 194 points of the original spiral data. Experiments were conducted with 10^2 , 10^3 , 10^4 , and 10^5 training samples which were drawn randomly from all 194 Gaussians. For comparison, Williamson (1996), used uniform distributions. Lee *et al.* (2004) fused a general regression neural network for prediction with fuzzy adaptive resonance theory and demonstrated better generalisation performance of their approach over Gaussian ARTMAP (Williamson 1996).

A spiral data set which was not directly derived from the original two-spiral data set was employed by Keuchel *et al.* (2003) to evaluate their method for unsupervised binary

partitioning in the area of computer vision. Their study also addressed the impact of geometric point distances on the partitioning result.

4. Experimental example

The aim of the following simulations is to show a state-of-the-art solution to the two-spiral task and to demonstrate how small variations of the training data can lead to classifiers with qualitatively different generalisation. The experiments employed LSSVMs with radial basis function kernels. This is similar to the way they were successfully used by Suykens and Vandewalle (1999a,b). In the LSSVM method, the typical inequality constraints of classical support vector machines (SVMs) and the associated quadratic optimisation procedure are replaced by equality constraints and an associated optimisation problem in the form of a system of linear equations. Figures 2 and 3 and all animations associated with the present study were generated using MatLab and the LSSVM toolbox by Pelckmans *et al.* (2002).

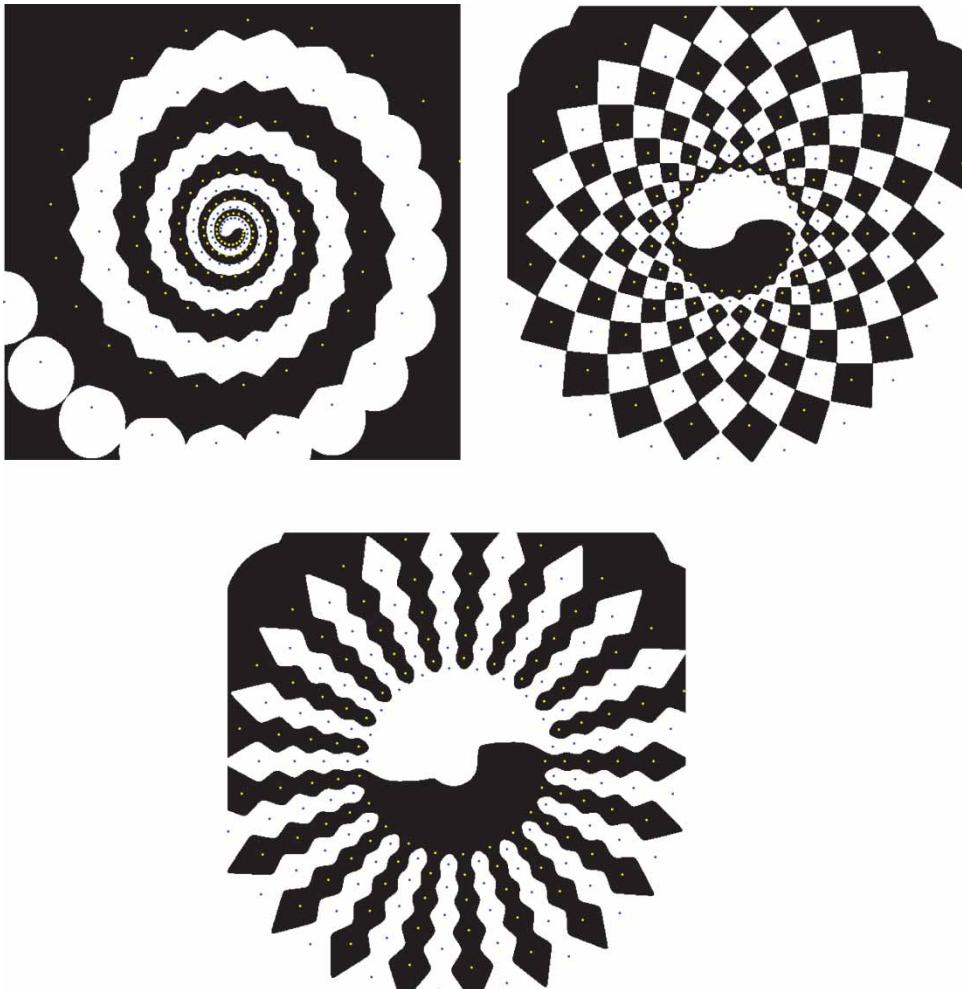


Figure 2. Changes in generalisation while varying the radial distance between the two spirals. This example uses data in configuration 2 and is explained in section 4.1.

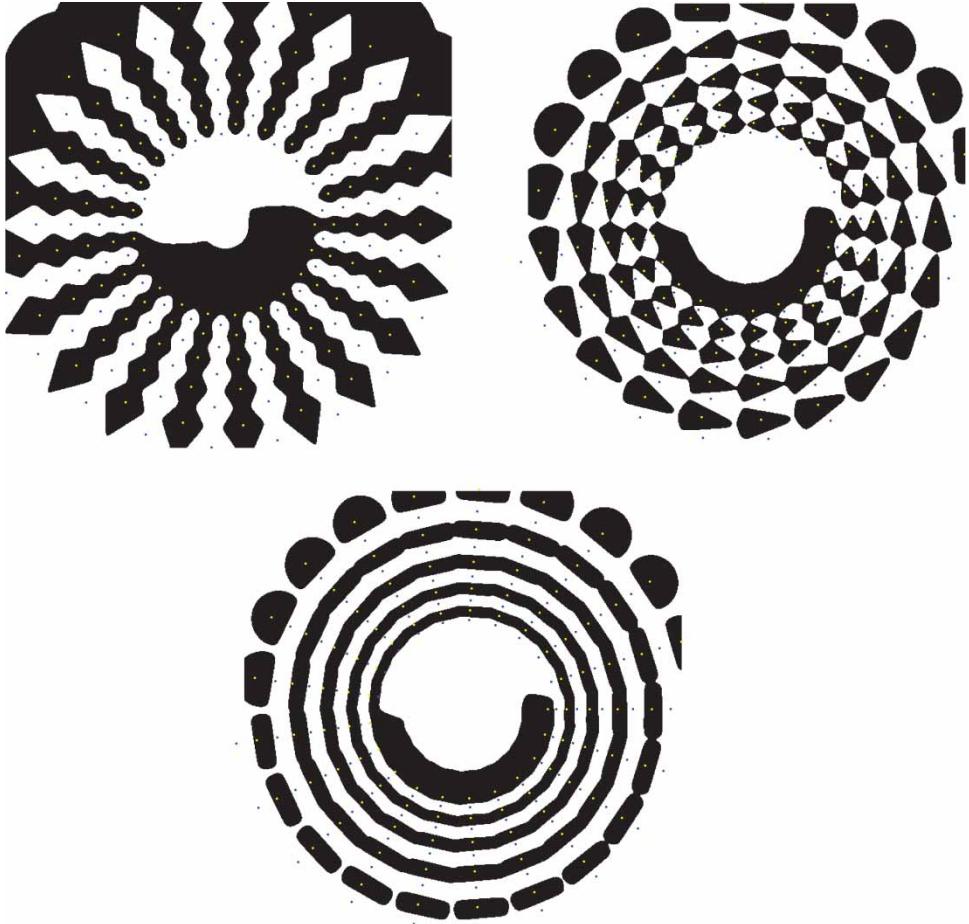


Figure 3. Changes in generalisation while rotating the two spirals relative to each other from configuration 2 into configuration 1 (section 4.2).

The spiral data for the simulation experiments were based on equation (5). This formula is a special case of the logarithmic spiral in equation (3), where the radial expansion of the spiral is controlled by the exponential parameter b and the linear expansion parameter is kept constant at $a = 1$.

$$\begin{pmatrix} x \\ y \end{pmatrix} = e^{b\phi} \cdot \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix} \quad (5)$$

For good visualisation it was decided, after several tests, to display all results using logarithmic spirals with five loops and 21 points per loop as shown in figures 2 and 3. Each of the displayed two-spiral data sets contained therefore 210 data points in total.

It is important to note that by small variation of the relative rotation of the two spirals two different configurations of the data set can be generated:

Configuration 1: The radial rays contain alternating dark and bright dots as shown in the bottom image of figure 3. This is essentially the same configuration of the two spirals as plotted in the original CMU data set (figure 1).

Configuration 2: The radial rays contain either dark or bright dots as shown, for example, in figure 2 and the top left image of figure 3.

For each of the two configurations, a series of simulation experiments was animated as a movie. The results are shown in figures 2 and 3. The associated movie files `dist.avi` and `rot.avi` are available at the weblinks <http://www.cs.newcastle.edu.au/~chalup/spirals/dist.avi> and <http://www.cs.newcastle.edu.au/~chalup/spirals/rot.avi>, respectively. Figures 2 and 3 show three frames (snapshots) corresponding to qualitatively different stages of the two animated simulation experiments. Each of the images/frames corresponds to one new training experiment and visualises the generalisation performance of an individually trained LSSVM classifier.

4.1 *Variation of the radial expansion of the spirals*

In this series of experiments, the two-spiral data were always kept in configuration 2 with equally spaced rays. Only the radial expansion of the spirals was modified in small increments for each frame of the movie such that the space between the two spirals became smaller step by step. At the start of the series of experiments (top left image of figure 2), the radial point distances were greater than the tangential point distances which forced the classifier to generalise towards spirals. At the final stage of the experiments (bottom image of figure 2), the radial distances were smaller than the tangential distances and it had become more efficient to separate the bright from the dark points using radial separator lines than spirals. At an intermediate stage (top right image of figure 2), the distances to neighbours in radial and tangential direction became approximately equal. Since spiral and ray decision boundaries exclude each other, this situation led to a diamond shaped generalisation pattern, where all but the most interior and exterior spiral points were separated from each other individually.

4.2 *Variation of the relative rotation of the spirals*

The second series of simulations started using the data points in configuration 2, where neighbour points of the same spiral in radial direction were closer to each other than neighbours in tangential direction. In accordance with the results of section 4.1, generalisation resulted in radial rays (top left image of figure 3). Figure 3 shows three snapshots from a series of experiments where the spirals were rotated relative to each other from configuration 2 to configuration 1 in small steps of $\pi/21$ radians. The top right image of figure 3 corresponds to an intermediate position and shows a diamond shaped solution similar to the one described in section 4.1. Configuration 1 allowed only generalisation with spiral shaped regions (bottom image of figure 3).

5. Summary and discussion

Solely based on studies which used the two-spiral task, it was possible to sketch a notable portion of the history of artificial neural networks up to date. Constructive methods and certain divide and conquer approaches were the two major initial streams of development triggered by the two-spiral task. Most of the developed methods show that approximate solutions are possible with combinations of linear separators or higher-order linear features. State-of-the-art learning methods which perform well on the two-spiral task include support vector machines and specifically tuned neural networks as outlined in section 3.1.4. One of the methods which can deliver very good results is the use of LSSVMs. These were applied in our example experiments.

Although most reviewed studies claimed that their method performed better than other previously published similar approaches, it should be acknowledged that the original paper by Lang and Witbrock (1988) had already achieved a lot, including an approximate solution using a non-standard MLP and evaluations on variations of the data.

An issue which became apparent through the survey was that studies on learning the two-spiral data typically did not provide enough experimental details to allow us to compile a precise global performance comparison of all studies. Many of the papers only provided statements based on visual inspection of the spirals' generalisation plots or some incomplete information on convergence rates. The latter was often expressed using 'numbers of epochs' which are hard to compare because the different methods' investment into a single 'epoch' has a high variance. Some studies addressed this issue by introducing notions such as 'connection crossings' which was defined as the number of multiply-accumulate steps required to propagate activation values forward, and error values backward through the network (Fahlman and Lebiere 1990, Treadgold and Gedeon 1997a,b). Several studies attempted to generate comparative performance tables which typically included only a very small number of alternative methods (*e.g.* Baum and Lang 1991, Carpenter *et al.* 1992, Fritzsche 1994, Han and Moraga 1995, Karayiannis 1996, Treadgold and Gedeon 1997a,b, Chaiyaratana and Zalzal 1998, Kamarthis and Pittner 1999, Lee and Fu 1999, Cheng and Zell 2000, Fan and Zhang 2000, Fu *et al.* 2001, Yang and Kao 2001, Ciocoiu 2002a, Ribeiro 2002, Thangavel and Kathirvalavakumar 2003, Du and Chen 2007). We refrained from summarising the latter information on epoch counts or classification accuracy in a global overview table because the non-standardised representation of results of the different studies could lead to false conclusions if taken out of their study's context.

The review revealed that the two-spiral task or its variations were frequently employed in pilot studies or studies with a non-experimental focus which were primarily interested in a quick and uncomplicated test or evaluation of their method. Here, its compact size and visual appeal made the two-spiral task one of the favourite candidates which turned out to be very useful to obtain indicative generalisation results by visual inspection and to facilitate communication between researchers. The number of studies which carefully tried to investigate the nature of the spirals data in systematic experiments was low.

Its strict regularity and the fact that slight variations of the two-spiral training data can lead to qualitatively different generalisation results could raise the question whether the two-spiral data is a valid benchmark data set for testing and comparing generalisation ability of machine-learning classifiers. It should be considered, for example, that the data is artificially generated and not iid[†]. The latter is a necessary condition for many results and concepts in machine-learning theory which address the topic of generalisation, such as PAC learning (Valiant 1984) and the VC-dimension (Vapnik 1979, 1995).

Concerns about the practice of using some artificially generated data sets for benchmarking have been stressed by several authors including Aha (1992) and Prechelt (1994). The latter, for example, argued that some algorithms may be able to solve artificial tasks very well, where a simple exact solution is known *a priori* while they might not be able to show the same performance on real world tasks. The survey in section 3.1.3 identified several studies which employed such a problem-specific bias, or preprocessed the data, for example, by a transformation into polar coordinates to simplify the two-spiral task. The problematic nature of this approach was also generally addressed by Denker *et al.* (1987) who claimed:

We prove that 'automatic learning will always succeed, given the right preprocessor', but we also show that this statement is grossly misleading since there is no automatic procedure for constructing the required preprocessor.

[†]The iid condition means that the training examples need to be generated by mutually independent random variables and the (marginal) probability distribution for each random variable is the same (Casella and Berger 2002).

The same argument could be applied to kernel selection for support vector machines, where not every kernel which works well for the two-spiral task would necessarily have to be a good kernel for general non-linear classification.

While the value of some of the studies with task-specific bias may be questionable, it should be acknowledged that the two-spiral task's regularity allows specific control and the opportunity to highlight and evaluate certain properties of a method.

Several studies employed variations of the spirals data which can provide additional insight. For example, noise addition was frequently applied to reduce regularity. However, again, often the associated articles did not provide sufficient details of the noise and the task to allow direct comparisons between different studies.

In summary, although strictly speaking there can be theoretical issues by using artificial data for benchmarking, the two-spiral task represents a compact test laboratory which has significantly contributed to the development of non-linear machine-learning classifiers.

6. Conclusion

The challenge of the two-spiral task inspired many researchers, studies, and new methods, and has had significant impact on machine-learning research over about 20 years. Its high popularity is partially due to its visual appeal and compact size. Given the large number of studies on the two-spiral task, it was possible to conduct a partial survey of central parts of the history of artificial neural networks solely based on studies which involve the two-spiral task either as their central task or in addition to some other data sets. An example experiment demonstrated the generalisation ability of a state-of-the-art solution of the two spiral task. The example experiment also showed that small variations of the two-spiral data can lead to qualitatively different generalisation results: spirals, rays or an intermediate diamond shaped solution.

Acknowledgements

The authors thank the anonymous reviewers for their useful comments. The first author is grateful to AIFB at the University of Karlsruhe in Germany and RIKEN BSI in Japan for providing an excellent work environment during sabbatical leave.

References

- D.W. Aha, Generalizing from case studies: A case study, in *Proceedings of the Ninth International Conference on Machine Learning*, D. Sleeman and P. Edwards, Eds, Morgan Kaufmann, San Mateo, CA, USA, 1992, pp. 1–10.
- J.R. Alvarez-Sanchez, Injecting knowledge into the solution of the two-spiral problem, *Neural Comput. and Appli.* 8, 265–272, 1999.
- M. Anthony and P.L. Bartlett, *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, Cambridge, UK, 1999.
- R.C. Archibald, The logarithmic spiral. *American Mathematical Monthly* 25, 189–193, 1918.
- N.E. Ayat, M. Cheriet, L. Remaki and C.Y. Suen, KMOD-A new support vector machine kernel with moderate decreasing for pattern recognition. Application to digit image recognition, in *Sixth International Conference on Document Analysis and Recognition (ICDAR'01)*. IEEE Computer Society, Los Alamitos, CA, USA, 2001, pp. 1215–1219.
- A. Baraldi and E. Alpaydin, Constructive feedforward ART clustering networks — part II. *IEEE Trans. Syst. Man Cybern. B*, 13, 662–677, 2002.
- P.L. Bartlett, The sample complexity of pattern classification with neural networks. The size of the weights is more important than the size of the network, *IEEE Trans. Inf. Theory* 44, 525–535, march 1998.
- E.B. Baum and K.J. Lang, Constructing hidden units using examples and queries, in *Advances in Neural Information Processing Systems (NIPS)*, P. Lippmann, J.E. Moody and S. Touretzky, Eds, Vol. 3, 1991, pp. 904–910.

- R.E. Blahut, *Digital Transmission of Information*, MA: Addison-Wesley, Reading, 1990, pp. 306–313.
- C.L. Blake and C.J. Merz, UCI repository of machine learning databases. Available on the internet at: <http://www.ics.uci.edu/~mllearn/MLRepository.html> (1998).
- A.P. Bradley and B.C. Lovell, *Inductive learning using multiscale classification*, in A.C. Tsoi and T. Downs, Eds. The University of Queensland, Brisbane, Australia, *Proceedings of the Fifth Australian Conference on Neural Networks, ACNN'94*, 1994, pp. 133–136.
- P.C. Bressloff and J.D. Cowan, Pattern formation, neural, in M.A. Arbib, (Ed), *The Handbook of Brain Theory and Neural Networks* (2nd edn), Cambridge, MA, USA: The MIT Press, 2003 pp. 859–863.
- P.C. Bressloff, J.D. Cowan, M. Golubitsky, P.J. Thomas and M.C. Wiener, Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex. *Philos. Trans. Royal Soc. London B Biolo. Sci.*, 356, 299–330, March 2001.
- J. Bruske and G. Sommer, Dynamic cell structure learns perfectly topology preserving map, *Neural Comput.* 7, 845–865, 1995.
- G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds and D.B. Rosen, Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Trans. Neural Net.* 3, 698–713, September 1992.
- G.A. Carpenter, S. Grossberg and J.H. Reynolds, A fuzzy ARTMAP nonparametric probability estimator for non-stationary pattern recognition problems, *IEEE Trans. Neural Netw.* 6, 1330–1336, November 1995.
- G. Casella and R.L. Berger, *Statistical Inference* (2nd edn). Duxbury Press, Pacific Grove, California, 2002.
- J.L. Castro, M. Delgado and C.J. Mantas, SEPARATE: a machine learning method based on semi-global partitions, *IEEE Trans. Neural Netw.*, 11, 710–720, May 2000.
- N. Chaiyaratana and A.M.S. Zalzal, Evolving hybrid RBF-MLP networks using combined genetic/unsupervised/supervised learning, in *UKACC International Conference on Control '98, 1–4 September 1998, University of Wales, Swansea, UK*, IEE Conference Publications No. 455. Institution of Electrical Engineers, London, UK, 1998, pp. 330–335.
- J. L. Chen and J. Y. Chang, Fuzzy perceptron neural networks for classifiers with numerical data and linguistic rules as inputs, *IEEE Trans. Fuzzy Syst.*, 8, 730–745, December 2000.
- K. Chen and D. Wang, Perceiving geometric patterns: From spirals to inside–outside relations, *IEEE Trans. Neural Netw.*, 12, 1084–1102, September 2001.
- K. Chen, L. Yang, Y. Xiang and H. Chi. A self-generating modular neural network architecture for supervised learning. *Neurocomputing*, 16, 33–48, 1997.
- G. Cheng, T. Liu, J. Han and Z. Wang, A. Towards growing self-organizing neural networks with fixed dimensionality. *Trans. Eng. Comp. Technol.*, 15, 268–272, October 2006.
- G. Cheng, Z. Song, J. Yang and R. Gao, B. On growing self-organizing neural networks without fixed dimensionality, in *International Conference on Computational Intelligence for Modelling, Control and Automation, 2006 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, *IEEE Comp. Soc.*, November 2006, pp. 164–168.
- G. Cheng and A. Zell, Externally growing cell structures for pattern classification, in *Proceedings of the Second ICSC Symposium on Neural Computation — NC'2000*, H. Bothe and R. Rojas, Eds, ICSC Academic Press, Canada/Switzerland, 2000, pp. 233–239.
- C.-C. Chiang and H.-C. Fu, 1994. A divide-and-conquer methodology for modular neural network design, in *IEEE International Conference on Neural Networks 1994 (ICNN'94)*. IEEE World Congress on Computational Intelligence, Vol. 1, 1994. pp. 119–124.
- J.-H. Chiang and P.-Y. Hao, Support vector learning mechanism for fuzzy rule-based modeling: A new approach, *IEEE Trans. Fuzzy Syst.* 12, 1–12, February 2004.
- H. Chua, J. Jia, L. Chen and Y. Gong, Solving the two-spiral problem through input data encoding. *Electroni. Lett.*, 31, 813–814, 1995.
- I.B. Ciocoiu, A modular feedforward network for solving classification problems, in *Proceedings of NATO Advanced Research Workshop on: Limitations and Future Trends in Neural Computation, LFTNC'01, Siena, Italy, 22–24 October, 2001*, M. Maggini (Ed.), pp. 69–72.
- I.B. Ciocoiu, Hybrid feedforward neural networks for solving classification problems, *Neural Process. Lett.*, 16, 81–91, 2002a.
- I.B. Ciocoiu, RBF networks training using a dual extended Kalman filter, *Neurocomputing*, 48, 609–622, 2002b.
- F.J. de Souza, M.M.R. Vellasco and M.A.C. Pacheco, Hierarchical neuro-fuzzy quadtree models. *Fuzzy Sets Syst.*, 130, 189–205, September 2002.
- J. Denker D. Schwartz, B. Wittner S. Solla, R. Howard, L. Jackel and J.J. Hopfield, Large automatic learning, rule extraction, and generalization, *Complex Syst.*, 1, 877–922, 1987.
- T. Denooux and R. Lengelle, Initializing backpropagation networks with prototypes, *Neural Netw.*, 6, 351–363, 1993.
- B. Dolin, M.G. Arenas, J.J. Merelo, Opposites attract: Complementary phenotype selection for crossover in genetic programming, in J.-J.M. Guervós, P. Adamidis, H.-G. Beyer, J.-L. Fernández-Villacanas and H.-P. Schwefel, Eds., *Parallel problem solving from nature — PPSN VII*, in *7th International Conference, Granada, Spain, 7-11, September, 2002*. Proceedings. Vol. 2439 of Lecture Notes in Computer Science (LNCS). Springer-Verlag, pp. 142–152.
- H. Du and Y.Q. Chen, 2007. Rectified nearest feature line segment for pattern classification, *Pattern Recognit.* 40, 1486–1497.

- S.E. Fahlman, An empirical study of learning speed in back-propagation networks. Computer Science Department Technical Report CMU-CS-88-162, Carnegie Mellon University, Pittsburgh PA, 1998a.
- S.E. Fahlman, Faster-learning variations on back-propagation: an empirical study, in *Proceedings 1988 Connectionist Models Summer School*, D.S. Touretzky, G.E. Hinton and T.J. Sejnowski, Eds, Morgan Kaufmann Publishers, Inc. San Mateo, CA: pp. 38–51. 1998b.
- S.E. Fahlman and C. Lebiere, The cascade-correlation learning architecture, in *Advances in Neural Information Processing Systems (NIPS)* D.S. Touretzky (Ed.), Vol. 2, 1990, pp. 524–532.
- S.E. Fahlman and C. Lebiere, The cascade-correlation learning architecture. Technical Report CMU-CS-90-100, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, August 1991.
- G. Fahner and R. Eckmiller, Structural adaption of parsimonious higher-order neural classifiers. *Neural Netw.*, 7, 279–289, 1994.
- W. Fan and L. Zhang, Applying SP-MLP to complex classification problems. *Pattern Recognit. Lett.*, 21, 9–19, 2000.
- B. Fritzke, Growing cell structures a self-organizing network for unsupervised and supervised learning, *Neural Netw.*, 7, 1441–1460, 1994.
- H.-C. Fu, Y.-P. Lee, C.-C. Chiang and H.-T. Pao, Divide-and-conquer learning and modular perceptron networks, *IEEE Trans. Neural Netw.*, 12, 250–263, March 2001.
- I. Gallese and J. Castellanos, Optimization of the kernel functions in a probabilistic neural network analyzing the local pattern distribution, *Neural Comput.*, 14, 1183–1194, 2002.
- M.S. Glassman, A network of localized linear discriminants, in *Advances in Neural Information Processing Systems (NIPS)* J.E. Moody, S.J. Hanson and R. Lippmann, Eds, Vol. 4 Morgan Kaufmann, pp. 1102–1109, 1992.
- S. Grossberg and L. Wyse, A neural network architecture for figure-ground separation of connected scenic figures, *Neural Netw.*, 4, 723–742, 1991.
- J. Han and C. Moraga, *Hybrid nets with variable parameters: a novel approach to fast learning under backpropagation*, in *INBS'95: Proceedings of the First International Symposium on Intelligence in Neural and Biological Systems*, IEEE Computer Society, Washington, DC, USA, 1995, pp. 72–75.
- P. Hilton, D. Holton and J. Pedersen, *Mathematical Reflections in a Room with many Mirrors*. Springer-Verlag, New York, 1997.
- M.J.J. Holt, Sequential learning of casaset network classifiers: constructing piecewise spherical decision boundaries, in *Proceedings of the International Conference of Artificial Neural Networks (ICANN'94)*, Sorrent M. Marinaro and P.G. Morasso, Eds, Springer-Verlag, Heidelberg, 1994a, pp. 1412–1415.
- M.J.J. Holt, A weight learning strategy for sequential construction of casaset network classifiers, in *Proceedings of the International Conference of Artificial Neural Networks (ICANN'94)*, Sorrent. M. Marinaro and P.G. Morasso, Eds, Springer-Verlag, Heidelberg, 1994b, pp. 1416–1419.
- I. Howard, Proposals for the study of anomalous perceptual schemata. *Perception*, 3, 497–513, 1974.
- D.-S. Huang and W.-B. Zhao, Determining the centers of radial basis probabilistic neural networks by recursive orthogonal least square algorithms, *Appl. Math. Comput.*, 162, 461–473, March 2005.
- J.-N. Hwang, S.-S. You, S.-R. Lay and I.-C. Jou, The cascade-correlation learning: a projection pursuit learning perspective, *IEEE Trans. Neural Netw.*, 7, 278–289, March 1996.
- J. Jia and H. Chua, Solving the two-spiral problem through input data representation, in *Proceedings of the IEEE International Conference on Neural Networks*, Vol. 1. 1995, pp. 132–135.
- H. Juille and J.B. Pollack, Co-evolving intertwined spirals, in *Evolutionary Programming V: Proceedings of the Fifth Annual Conference on Evolutionary Programming*, L.J. Fogel, P.J. Angeline and T. Baeck, Eds, The MIT Press, Cambridge, MA, USA, 1996, pp. 461–467.
- S.V. Kamarthi and S. Pittner, Accelerating neural network training using weight extrapolations, *Neural Netw.*, 12, 1285–1299, 1999.
- W. Kaminski and P. Strumillo, Kernel orthonormalization in radial basis function neural networks, *IEEE Trans. Neural Netw.*, 8, 1177–1183, September 1997.
- N.B. Karayiannis, Hybrid learning schemes for fast training of feed-forward neural networks, *Math. Comp. Simul.*, 41, 13–28, 1996.
- J. Keuchel, C. Schnörr, C. Schellewald and D. Cremers, Binary partitioning, perceptual grouping, and restoration with semidefinite programming. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25, 1364–1379, November 2003.
- J.R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MA, USA: The MIT Press, Cambridge, 1992.
- L.I. Kuncheva, Initializing of an RBF network by a genetic algorithm, *Neurocomputing*, 14, 273–288, 1997.
- J.J.T. Lahnajärvi, M.I. Lehtokangas and J.P.P. Saarinen, Evaluation of constructive neural networks with cascaded architectures, *Neurocomputing*, 48, 573–607, October 2002.
- K.J. Lang and M.J. Witbrock, Learning to tell two spirals apart, in *Proceedings 1988 Connectionist Models Summer School*, Morgan Kaufmann, Los Altos, CA: D. Touretzky, G. Hinton and T. Sejnowski, Eds, 1988, pp. 52–59.
- E.W.M. Lee, C.P. Lim, R.K.K. Yuen and S.M. Lo, A hybrid neural network model for noisy data regression. *IEEE Trans. Syst. Man Cyb. B: Cyb.*, 34, 951–960, April 2004.
- Y.-P. Lee and H.-C. Fu, Weight estimation for the learning of modular perceptron networks, in *Neural Networks for Signal Processing IX*, Proceedings of the 1999 IEEE Signal Processing Society Workshop, 1999, pp. 103–111.
- M. Lehtokangas, Determining the number of centroids for CMLP network, *Neural Netw.*, 13, 525–531, 2000.
- M. Lehtokangas and J. Saarinen, Centroid based multilayer perceptron networks, *Neural Process. Lett.* 7, 101–106, April 1998a.

- M. Lehtokangas and J. Saarinen, Weight initialization with reference patterns, *Neurocomputing*, 20, 265–278, August 1998b.
- B.C. Lovell and A.P. Bradley, The multiscale classifier, *IEEE Trans. Patt. Anal. Mach. Intell.*, 18, 124–137, February 1996.
- B.-L. Lu and M. Ito, Task decomposition and module combination based on class relations: A modular neural network for pattern classification, *IEEE Trans. Neural Netw.*, 10, 1244–1256, September 1999.
- W.S. McCulloch and W. Pitts, A logical calculus of ideas immanent in nervous activity, *Bull. Math. Biophys.*, 5, 115–133, 1943.
- M.L. Minsky and S. Papert, *Perceptrons: an introduction to Computational Geometry*, Cambridge, MA, USA: The MIT Press, 1969.
- M.L. Minsky and S. Papert, *Perceptrons: an introduction to Computational Geometry, (expanded edition)*, Cambridge, MA, USA: The MIT Press, 1988.
- E. Mizutani and S.E. Dreyfus, MLP's hidden node saturations and insensitivity to initial weights in two classification benchmark problems: parity and two-spirals, in *Proceedings of the 2002 International Joint Conference on Neural Networks (IJCNN '02)*, Vol. 3, IEEE Press, 2002, pp. 2831–2836.
- S. Osowski, P. Bojarczak and M. Stodolski, Fast second order learning algorithm for feedforward multilayer neural networks and its applications, *Neural Netw.*, 9, 1583–1596, 1996.
- G. Patanè and M. Russo, Fully automatic clustering system, *IEEE Trans. Neural Netw.*, 13, 1285–1298, November 2002.
- K. Pelckmans, J.A.K. Suykens, T. Van Gestel, J. De Brabanter, L. Lukas, B. Hamers, B. De Moor and J. Vandewalle, LS-SVM lab: a Matlab/C toolbox for least squares support vector machines. Internal Report 02-44, ESAT-SISTA, K.U. Leuven, Belgium, Leuven, Belgium, 2002.
- M.A. Potter and K.A. De Jong, Cooperative coevolution: an architecture for evolving coadapted subcomponents, *Evol. Comp.*, 8, 1–29, 2000.
- L. Prechelt, 1994. PROBEN 1 – a set of benchmarks and benchmarking rules for neural network training algorithms. Technical report 21, Universität Karlsruhe; Institut für Programmstrukturen und Datenorganisation, Germany.
- J.R. Quinlan, *C4.5: Programs for Machine Learning*, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.
- B. Ribeiro, On data based learning using support vector clustering, in *Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'02)*, L. Wang, J.C. Rajapakse, K. Fukushima, S.-Y. Lee and X. Yao, Eds, Vol. 5, 2002, pp. 2516–2521.
- S. Ridella, S. Rovetta and R. Zunino, Circular backpropagation networks for classification. *IEEE Trans. Neural Netw.*, 8, 84–97, January 1997.
- S. Ridella, S. Rovetta and R. Zunino, Representation and generalization properties of class-entropy networks. *IEEE Trans. Neural Netw.*, 10, 31–47, January 1999.
- F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain, *Psychol. Rev.* 65, 386–408, 1958.
- R. Setiono, A neural network construction algorithm which maximizes the likelihood function, *Connection Sci.*, 7, 147–166, 1995.
- Y. Shang and B.W. Wah, Global optimization for neural network training. *IEEE Comp.*, 29, 45–54, 1996.
- T.R. Shultz and J.L. Elman, Analysing cross connected networks, in *Advances in Neural Information Processing Systems (NIPS)*, Vol. 6, 1994, pp. 1117–1124.
- T.R. Shultz, Y. Oshima-Takane and Y. Takane, Analysis of unstandardized contributions in cross connected networks, in *Advances in Neural Information Processing Systems (NIPS)*, Vol. 7, 1995 pp. 601–608.
- S. Singh, 2D spiral pattern recognition with probabilistic measures, *Patt. Recogn. Lett.*, 19, 141–147, 1998.
- S. Singh, Neural learning of spiral structures, in *Proceedings International Conference on Advances in Pattern Recognition (ICAPR'98)*, Singh, S. (Ed.). Plymouth UK, Springer-Verlag, 1998b, pp. 226–231.
- S. Singh, Quantifying structural time varying changes in helical data, *Neural Comp. Appl.*, 10, 148–154, 2001.
- F. Smieja, The pandemonium systems of reflective agents, *IEEE Trans. Neural Netw.*, 7, 97–106, January 1996.
- M. Solazzi and A. Uncini, Artificial neural networks with adaptive multidimensional spline activation functions, in *IJCNN-2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, S.-I. Amari, C.L. Giles, M. Gori and V. Piuri, Eds, *Como, Italy, 24–27 July, 2000*, Vol. 3, IEEE Computer Society, pp. 471–476.
- J.M. Sopena, E. Romero and R. Alqazar, Neural networks with periodic and monotonic activation functions: a comparative study in classification problems, in *Proceedings of the 9th International Conference on Artificial Neural Networks (ICANN'99)*, D. Willshaw and A. Murray, Eds, University of Edinburgh. IEE Conference Publication No. 470, 1999, pp. 323–328.
- L. Su and S.-U. Guan, Two-dimensional extensions of cascade correlation networks, in *The Fourth International Conference on High-Performance Computing in the Asia-Pacific Region (HPC-ASIA 2000)*, Vol. 1, IEEE Computer Society, 2000, pp. 138–141.
- C.T. Sun and J.S. Jang, A neuro-fuzzy classifier and its applications, in *Proceedings of the Second IEEE International Conference on Fuzzy Systems, San Francisco, CA, USA, March 1993*, Vol. 1, 1993, pp. 94–98.
- J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle, *Least Squares Support Vector Machines*, Singapore: World Scientific Pub. Co., and 2002.
- J.A.K. Suykens and J. Vandewalle, Least squares support vector machine classifiers, *Neural Process. Lett.*, 9, 293–300, 1999a.

- J.A.K. Suykens and J. Vandewalle, Training multilayer perceptron classifiers based on a modified support vector method. *IEEE Trans. Neural Netw.*, 10, 907–911, July 1999b.
- C.-C. Teng and B.W. Wah, Automated learning for reducing the configuration of a feed-forward neural network, *IEEE Trans. Neural Netw.*, 7, 1072–1085, September 1996.
- P. Thangavel and T. Kathirvalavakumar, Simultaneous perturbation for single hidden layer networks—cascade learning, *Neurocomp.*, 50, 193–209, January 2003.
- W. Ting and Y. Sugai, A wavelet neural network for January 2003. the approximation of nonlinear multi-variable function, in *1999 IEEE International Conference on Systems, Man, and Cybernetics, IEEE SMC '99 Conference Proceedings*, October 1999. Vol. 3, pp. 378–383.
- D. Tomandl and A. Schober, A modified general regression neural network (MGRNN) with new, efficient training algorithms as a robust ‘black box’-tool for data analysis, *Neural Netw.*, 14, 1023–1034, 2001.
- N.K. Treadgold and T.D. Gedeon, A cascade network algorithm employing progressive RPROP. in *Biological and Artificial Computation: from Neuroscience to Technology, International Work-Conference on Artificial and Natural Neural Networks, IWANN '97, Lanzarote, Canary Islands, Spain, 4–6, June*, J. Mira, R. Moreno-Díaz and J. Cabestany, Eds, 1997, Proceedings, Vol. 1240 of Lecture Notes in Computer Science (LNCS). Springer-Verlag, pp. 733–742.
- N.K. Treadgold and T.D. Gedeon, Extending and benchmarking the CasPer algorithm, in *Advanced Topics in Artificial Intelligence, 10th Australian Joint Conference on Artificial Intelligence, AI'97, Perth, Australia*, A. Satter, (Ed.), Vol. 1342 of Lecture Notes in Computer Science (LNCS). Springer-Verlag, 1997b, pp. 398–406.
- N.K. Treadgold and T.D. Gedeon, Exploring constructive cascade networks, *IEEE Trans. Neural Netw.*, 10, 1335–1350, November 1999.
- L.G. Valiant, A theory of the learnable. *Comm. ACM*, 27, 1134–1142, November 1994.
- V.N. Vapnik, *Estimation of Dependences Based on Empirical Data* [in Russian]. Nauka, Moscow, English translation: New York: Springer-Verlag, 1979.
- V.N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.
- B.W. Wah and M. Qian, Constrained formulations for neural network training and their applications to solve the two-spiral problem, In *Proceedings of the 5th International Conference on Computer Science and Informatics and 5th Joint Conference on Information Sciences*, Vol. 5, 2000, pp. 598–601.
- M.K. Weir and L.H. Chen, Neural training and generalisation of sequences using continuous temporal structure, in *1991 IEEE International Joint Conference on Neural Networks*, Vol. 3. IEEE, 1991 pp. 2027–2032.
- E.W. Weisstein, Eric Weisstein’s world of mathematics—a Wolfram web resource. Available online at: <http://mathworld.wolfram.com> (2003).
- M. White, T. Sejnowski, C. Rosenberg, N. Qian, R.P. Gorman, A. Wieland, D. Deterding, M. Niranjani and T. Robinson, Bench: CMU neural networks benchmark collection. Available at: <http://www-2.cs.cmu.edu/afs/cs/project/airepository/ai/areas/neural/bench/cmu/0.html> (1995).
- D. Whitley and N. Karunanithi, Generalization in feed-forward neural networks, in *Proceedings of the International Joint Conference on Neural Networks (IJCNN'91)*. Vol. II, IEEE Press, 1991, pp. 77–82.
- J.R. Williamson, Gaussian ARTMAP: a neural network for fast incremental learning of noisy multidimensional maps. *Neural Netw.* 9, 881–897, 1996.
- Q. Xiong, K. Hirasawa, J. Hu and J. Murata, Comparative study between functions distributed network and ordinary neural network, in *IEEE International Conference on Systems, Man, and Cybernetics*, Vol. 3, 2001 pp. 1548–1553.
- Q. Xiong, K. Hirasawa, J. Hu and J. Murata, A functions localized neural network with branch gates, *Neural Netw.*, 16, 1461–1481, 2003.
- J.-M. Yang and C.-Y. Kao, A robust evolutionary algorithm for training neural networks, *Neural Comput. Appl.*, 10, 214–230, 2001.
- S. Young and T. Downs, CARVE—a constructive algorithm for real-valued examples. *IEEE Trans. Neural Netw.*, 9, 1180–1190, November 1998.
- C.-C. Yu and Y.-C. Tang, B.-D. Liu, An adaptive activation function for multilayer feedforward neural network. in *TENCON'02. Proceedings. 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering*, Vol. 1, 2002 pp. 645–650.
- L. Zhang and B. Zhang, A geometrical representation of McCulloch–Pitts neural model and its applications, *IEEE Trans. Neural Netw.*, 10, 925–929, July 1999.

