ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 165-176

Neural Networks Approaches in Medicine – a Review of Actual Developments

Thomas Villmann
University Leipzig
Clinic for Psychotherapy and Psychosomatic Medicine
D–04107 Leipzig, Karl–Tauchnitz–Str.25, Germany
email: villmann@informatik.uni-leipzig.de

Abstract. The utilization of neural network approaches have been rapidly increasing also in the area of medicine. Thereby both biomedical modelling as well as data analysis are objects of neural network applications. On the one hand side neural networks give the possibility for understanding of brain and cognitition processes. On the other hand the power of neural networks for data analysis, data visualization and knowledge discovery is used. In the following we will reflect some of the actual developments and problems when neural networks are applied in medical area.

1. Why neural networks in medicine?

The number of neural network applications in medical area is drastically increasing in the last years. Thereby, neural network approaches can be viewed from two different basic points: On the one hand side, brain models and cognitive functions are described by network models and architectures. Examples are the concept of spiking neurons, the ART architecture, the Hopfield network, etc. (for an overview we refer to [16]). A widely ranged class of neural models for description of cortical phenomena is the concept of neural maps [2],[22],[23],[44]. Especially, cortical brain functions were extensively considered in the last years and have lead to a better understanding of organization processes in the sensomotorical area [32] as well as in the visual cortex [31],[5] or in the auditive cortex [38].

On the other hand neural networks can be taken as technical tools for data analysis, knowledge discovery and data compression, data visualization, time series prediction, process control and other. Thereby, the data occurring in medical applications are often characterized by at least one of the following properties which often cause difficulties:

- small data sets
- non-linear and high-dimensional data sets
- large noise without any clearly reproducible cause
- large data in image processing
- scaling and choice of metrics in multidimensional data

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 165-176

- categoric and non-metric data

Because of these reasons traditional approaches like conventional statistics and data analysis often fail. Artificial neural networks can help to overcome some of the above mentioned difficulties in medical applications. However, intelligent data analysis often requires so-called hybrid approaches combining several data analysis techniques which may include also other innovative methods like Fuzzy-approaches, stochastic optimization, genetic and evolutionary algorithms. In general, the problem of analysis of non-metric data frequently occurs in social science and medical applications [13, 36].

Some of interesting applications which have also lead to interesting new developments in the neural network or algorithm design are the in the area of image processing [45], time series analysis and prediction [21], time series classification [26], neuro-fuzzy approaches [10] and other, some of them can be found in this volume. In the following we will demonstrate how neural networks may be successfully applied for immediately medical assessment if used as a data visualizing tool. A second exemplary application concerns the clustering of categoric data using a hybrid approach based on an evolutionary algorithm including a neural dynamic.

2. Exemplary Applications

2.1. Visualization of physiological parameters for easy clinical assessment using extended SOM

The therapeutical process in psychotherapy usually is organized in a sequence of daily single therapy sessions. During the therapy and, especially during the single sessions, the emotional feeling can vary in dependence on the actual situation (emotional excitements, as the result of the therapeutical discussion) which can be observed by individual asses by the therapist. Beside the individual observations there exist a large pool of instruments to judge the emotional situation of both the patient and the therapist which consist in a spectrum of several questionnaires [19, 43]. On the other hand, it is generally known that emotions influence physiological parameters as for instance heart rate, respiration rate, muscle tension and electrodermal conductivity (skin conductance) which can end for negative cases in physiological complaints, called psychosomatic symptoms, which are under treatment in our clinic. Vice versa, the therapy process also influences the physiological parameters. The *parallel* observation of such psycho-physiological data during therapy sessions together with a psychological analysis should lead to a better understanding of the psycho-physiological processes. For this purpose a suitable tool is needed for an easy parallel assessment.

In this initial investigation the therapy of one patient was considered containing $t_{\max} = 37$ single sessions of approximately 45 Min.. During the several sessions for both the patient and, additionally, for the therapist the following parameters were simultaneously obtained: *heart rate,muscle tension (by surface electromyogram), skin conductance level (SCL), skin conduction reaction (SCR)*. The value are determined as averages over time intervals of $30s$. Here we are interested in the variation of the parameters. Hence, we investigate only the difference of the parameter for a certain time point in comparison to

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 165-176

the previous one. In this way we obtain 3105 data vectors for both the patient and the therapist. Hence, we have $n_{\max} = 6210$ data vectors for neural network learning. Subsequently, the data were normalized such that for the resulted vectors $\mathbf{v} \in \mathcal{V} \subseteq \Re^4$ we have for all components j: $\sum_{i=1}^{n_{\max}} v_j = 1$, i.e., all measured parameters are assumed to be of the same importance.

The data serve as input into a *Growing Self-Organizing Map (GSOM)* [9] which is an extension of the usual Self-Organizing Map (SOM) [23]. In general, neural maps project data from some (possibly high-dimensional) input space $\mathcal{V} \subseteq \Re^{D_\mathcal{V}}$ onto a position in the (usually hypercubical) neuron lattice $\mathcal{A}$ in a topographic manner. For this purpose, during data driven learning, the algorithm distributes pointers $\mathbf{W} = \{\mathbf{w_r}\}$ of the neurons $\mathbf{r} \in \mathcal{A}$ in the input space according to the input distribution $\mathcal{P}(\mathcal{V})$ such that

$$\mathcal{P}(\mathcal{V}) \sim \mathcal{P}(\mathbf{W})^\alpha \tag{1}$$

with $\alpha = \frac{2}{3}$ holds [33],[24].[1] The exponent α is called the magnification factor. Thereby topographic mapping means a topology preserving projection, i.e. a continuous change of a parameter of the input data should lead to a continuous change of the position of a localized excitation in the neural map.[2] Because of the lack of space we refer to [23],[32],[40] for detailed considerations. The GSOM not only adapts the neural pointers, additionally the lattice structure is adapted remaining a hypercube structure [9] and hence, improves the degree of topology preservation. In general, a better topology preservation leads to a better accuracy of the map. Several approaches were developed to judge the degree of topology preservation for a given map [7]. A robust tool is the topographic product P [8] which comprise the information about violations of the topology preservation in a single value. P can take positive and negative values indicating that the dimensionality of the grid $\mathcal{A}$ is too low dimensional or too high, respectively, for representation of the given data. If one obtains P approximately zero the shape of the lattice matches the data distribution. For a detailed introduction and analysis of the properties we refer to [8] and [40].

In this way it can be understood as a *nonlinear principle component analysis* [39]. A further extension of the basic SOM, also included in our application, concerns the magnification : the usual SOM distributes the pointers $\mathbf{W}$ according (1) Bauer et al. in [6] introduced a local learning parameter $\epsilon_\mathbf{r}$ with $\langle \epsilon_\mathbf{r} \rangle \propto \mathcal{P}(\mathcal{V})^m$ which finally leads to a relation $\mathcal{P}(\mathcal{V}) \sim \mathcal{P}(\mathbf{W})^{\alpha'}$ with $\alpha' = \alpha(m+1)$ and, hence, allows a magnification control. Especially, one can achieve a resolution $\alpha' = 1$ which maximizes the mutual information (corresponding to a maximization of the entropy) [27].

We applied the GSOM with included magnification control to the above therapy data and result a $7 \times 5 \times 7$–lattice for SOM. The topology preservation assessment by a special kind of the topographic product [37] yields a value $\tilde{P} = 0.007$ referring to a good topology preservation. The magnification control scheme have lead to a improvement of the entropy: without control the entropy was 94.8% of the maximal possible entropy [42] value in comparison to 96.4% using the control scheme.

Because of resulting a $3d$-lattice as the final shape for the grid $\mathcal{A}$ we can interpret the neuron grid positions as colors in the *three-dimensional* colorspace

[1] This result is valid for the one-dimensional case and higher dimenional ones which separate.

[2] In this way the SOM determines the non-linear principle components of the data.

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 165-176

as the intensity of the colors red, green and blue in the well known RGB-model. In this way we are able to code each data vector as a color pixel by application of the map $\Psi_{\mathcal{V}\to\mathcal{A}}$ realized by the trained network. After this an easy visual interpretation is possible by the therapists is possible as depicted in Fig. 1:

In Fig. 1 we plotted for each time point in all sessions the mapped original data according to the above outlined color coding. For example, we observe a

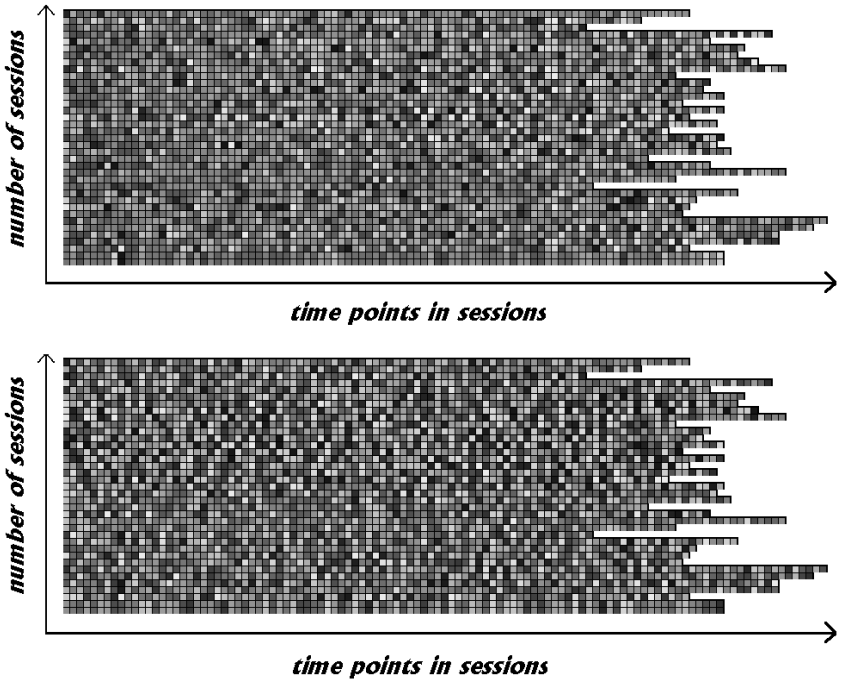

Figure 1: Color description of the variation of the physiological parameters for the patient (above) and the therapist (below) for the complete therapy for each single session. Each colored pixel code a characteristic pattern of change. Global color texture changes correspond to variations in the therapy (see text for details). Single colors can be related to charteristic patterns. The color representation is resulted from a GSOM generated $7 \times 5 \times 7$ SOM-lattice including a magnification control scheme for maximizing the entropy. (Regrettably, only a greyscale image can be provided in the printed paper. A color image is available on request from the author.)

clear change in the distribution of colors after the 20th session: for the therapist picture the occurrence yellow typed colors is increasing, in the patient picture in the same area the frequency of blue colored pixels is increased. Now, the following interpretation (in the sense of psychotherapy research) can be done: this change in the color distribution *is related to a change in the treatment concept regarding the therapeutical interactions*. An second alteration in

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 165-176

the color distribution of the therapist picture can be recognized for the last 5 sessions which are indicated by the detachment phase of the patient from the therapist which plays an important role in the psychotherapy process. This phase is indicated by a drastic change of the patient-therapist-relation. On the other hand, investigating the initial phases of therapy sessions we observe that the frequency of blue-green colors for the patient decreases. We have found that these colors correspond to a decreasing heart rate. This is in agreement to the assessment by therapist where is stated that in the first sessions there was a big emotional pressure and expectation from the patient in the beginning phase which is lost during the following ones.

Moreover, it is possible to assign characteristic patterns of parameter changes to specific colors by deeper analysis. For instance, as mentioned above, we found that green-blue colors are related to a decreasing heart rate whereas red color indicates an increase. Yellow-green colors symbolize increasing muscle tensions. The rise of electrodermal conductivity is coded by magenta-like colors. However, for a detailed considerations further investigations are necessary.

In conclusion, we can say that the demonstrated visualization technique of physiological parameters during psychotherapy sessions can help therapist to understand the psycho-physiological feeling of the patient.

2.2. Clustering of non-metric data using evolutionary algorithms

The investigation and analysis of non-metric data frequently occurs in medical application and social sciences. However, these problems are difficult to capture. One possibly way to handle such data is the application of *Genetic* or *Evolutionary Algorithms* (EAs). EAs are a biologically motivated stochastic iterative optimization method. In EAs the manipulation of objects, which are called individuals $\mathbf{s} \in \mathbf{\Pi}$, is separated from their evaluation by a fitness (objective) function F. The set $\mathbf{\Pi}$ is called population. There exist two basic manipulation operators: mutation as random change of parameters and the crossover as merging of two individuals. Usually the μ individuals of $\mathbf{\Pi}$ generate λ new ones with $\lambda > \mu$. After these manipulations F judges how proper the individuals fulfil the considered task.[3] If in one time step all individuals have gone under manipulation and evaluation, from these the *selection operator* extracts a new generation of the population for the next iteration step. For a detailed overview we refer to [3] or [29]. The basic advantage is that in EAs the manipulation of the object is strictly separated from their evaluation and, hence, a larger set of manipulation techniques are available.

In the following we demonstrate the application of EAs for clustering of categoric data in psychology research. At the same time we suggest some extensions of the basic EA schemes to improve the performance which are an *advanced selection strategy* and a *multiple population approach with migration* inspired by the learning dynamic in neural maps. Furthermore, we emphasize the aspect of a proper choice of the fitness function.

Modern psychology uses all the standard methods of mathematical statistics to extract relevant features, structural information and other data obtained from several therapeutical approaches. One of the mostly used method for

[3] Thereby, the fitness function may contain explicit expert knowledge regarding to the optimization task which may be difficult to code otherwise.

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 165-176

acquisition of structures of interpersonal relationships in the area of psycho–dynamic psychotherapy research is the method of the '*Core Conflictual Relationship Theme*' (CCRT) developed by LUBORSKY, [28]. The method investigates short stories about relationships, so–called *relationship–episodes*, which are often reported by the patients in their therapeutical sessions. In each of these episodes the components *wish of the subject*, *response of the object* and *response of the subject* were encoded which are used to perform the CCRT using a system $\mathcal{S}^w$ of $s^w_{\max} = 34$ so-called *standard categories* S^w_j to classify the wishes [12]. Each standard category describes an aspect in verbal manner. However, the categories are often correlated in meaning because of they are describing not only contrary but also the same or similar psychological topics. Therefore, they are collected in a set $\mathcal{C}^w$ of $c^w_{\max} = 8$ *clusters* C^w_k [4] which are then used in further considerations instead of the standard categories. Analogously, we have $s^{ro}_{\max} = 30$ categories $S^{ro}_j \in \mathcal{S}^{ro}$ for encoding the response of both the object and subject [12] which are collected in $c^{ro}_{\max} = 8$ clusters $C^{ro}_k \in \mathcal{C}^{ro}$ and $c^{ro}_{\max} = 8$ clusters $C^{rs}_k \in \mathcal{C}^{rs}$, respectively. However, the clusters also are still correlated again. Furthermore, several considerations have shown that *the used scheme of assignment still leads to unsatisfactory reliability rates* and interpretation problems in investigations based on this scheme. For further and more detailed critical remarks we refer to [1].

For solving the re-clustering problem of standard categories at first we determined the similarities between the standard categories S^w_i, S^w_j on the basis of their (symmetric) conditional probabilities $p^w(i \mid j)$ (in analogy $p^{ro}(i \mid j)$, $p^{rs}(i \mid j)$ for the responses of object and subject). Now, to perform the clustering, it has to be done as a clustering of proximity data. This problem we can apply the approach of vector quantization of proximity data introduced by HOFMANN & BUHMANN [18] and GREAPEL & OBERMAYER [15]:

$$\mathcal{H} = \frac{1}{2} \sum_{i=1}^{M} \sum_{k=1}^{M} \frac{D_{i,k}}{M} \left(\sum_{v=1}^{c_{\max}} \frac{u_{i,v} u_{k,v}}{p_v} - 1 \right) \tag{2}$$

with $p_v = \sum_{i=1}^{M} u_{i,v}/M$ the normalized percentage of the data in that cluster and M the number of data. The values $D_{i,k}$ are the *dissimilarities* between the data points which we can define as difference of the conditional probabilities from the unit. $\mathcal{H}$ serves as a cost function and has to be minimized. As pointed out in [18] $\mathcal{H}$ is independent on non-symmetric dissimilarities and permutations of the cluster indices. Moreover, the cluster assignments $u_{i,v}$ are distributed according to the Gibbs distribution

$$\mathbf{P}(\mathcal{H}) = \exp\left(-\frac{\mathcal{H} - \mathbf{F}(\mathcal{H})}{T} \right) \tag{3}$$

whereby T plays the role of a temperature, and $\mathbf{F}(\mathcal{H})$ is the free energy. Yet, because of the statistical dependence of the assignments the Gibbs distribution $\mathbf{P}(\mathcal{H})$ can not be exactly rewritten in a factorized form, however a mean field approximation and corrections to the assignment correlations has been derived [18]. Yet, $\mathcal{H}$ requires a *priori known value* $c_{\max}$, which is, in fact, a hard restriction and difficult to overcome [11]. For this purpose, the so-called *partition entropy* was introduced which *explicitly takes the number of clusters*

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 165-176

into account [20]:

$$\mathcal{E}_{\mathcal{P}}\left(c_{\max}\right) = -\frac{\frac{1}{M\log(c_{\max})}\sum_{k=1}^{M}\sum_{i=1}^{c_{\max}} u_{k,i}\log\left(u_{k,i}\right)}{1-\frac{c_{\max}}{M}} \quad . \tag{4}$$

$\mathcal{E}_{\mathcal{P}}\left(c_{\max}\right)$ is often applied in medical applications [13]. Yet, these approaches require metric spaces as data in contrast to the topographic vector quantizer. Therefore, if we combining the partition entropy $\mathcal{E}_{\mathcal{P}}\left(c_{\max}\right)$ (4) and $\mathcal{H}$ (2) we obtain as new fitness function:

$$\mathcal{F} = \frac{\frac{1}{M\ln(c_{\max})}\mathcal{H}}{1-\frac{c_{\max}}{M}} \tag{5}$$

which explicitly takes the number of clusters into account and is also applicable to non-metric data. Hence, $\mathcal{F}$ incorporates both demanded features. As it shown in [15] it is possible to minimize $\mathcal{H}$ by simulated annealing, and, hence this approach could be applied to optimize $\mathcal{F}$.

Yet, it is difficult to generate a careful annealing scheme, because the of the large number of local minima. A possibility for convergence improvement is the application of evolutionary algorithms to optimize $\mathcal{H}$ or $\mathcal{F}$. For the application of EAs for clustering as presented below it is useful to interpret clustering as a *partitioning problem* under constraints. Thereby, a *partitioning* of a nonempty set $\mathcal{S}$ related to a nonempty set $\mathcal{C}$ is an unique and surjective mapping

$$\Phi : \mathcal{S} \to \mathcal{C} \quad . \tag{6}$$

Then a *partition* Ψ_Φ of $\mathcal{S}$ with respect to the partitioning Φ is given by $\Psi_\Phi = \left\{\Phi^{-1}\left(C\right) \mid C \in \mathrm{cod}\left(\Phi\right)\right\}$ whereby $\mathrm{cod}\left(\Phi\right)$ is the range of Φ. These constraints are collected in a fitness function which may be identified as the above $\mathcal{F}$.

For solving the partitioning task each individual in a generation of an EA describes a certain partition. In the present paper we assume $\mathcal{S}$ to be discrete containing $s_{\max}$ elements. Then we can take each individual as a string of length $s_{\max}$ the components of which contain the cluster index onto which the respective component has to be mapped. Mutation of an individual is defined as a random change of the mapping for a randomly selected individual component, and the crossover is the usual two-point crossing.

Several approaches were developed to improve the above basic EA-scheme. Especially, multiple subpopulation approaches are widely considered [14, 30]. Thereby the basic population $\mathbf{\Pi}$ is divided into subpopulations Π_i which have more or less communication during the evolution which may be realized as migration. To improve the performance of the basic subpopulation approach we arranged the subpopulations Π_i in a topological order usually chosen to be a regular lattice, for instance a ring or a quadratic lattice [4]. Between these subpopulations a *migration scheme* was introduced which is adapted from the SOM-learning: a visit (migration) from individuals between neighboring subpopulations is allowed for a short time *regarding to the topological order* Ω in $\mathbf{\Pi}$ and with respect to a time-dependent neighborhood function

$$h_{i^*}\left(t,k\right) = \left(1.0-\epsilon_h\right)\cdot\exp\left(-\frac{r_{i^*,k}}{2\left(\sigma_h\left(t\right)\right)^2}\right)+\epsilon_h \tag{7}$$

[4] In general, other arrangements are also admissible. Then the lattice can be defined by a connection matrix which describes the neighborhood relations.

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 165-176

with a small positive number ϵ_h [41]. During the evaluation of a certain subpopulation Π_{i^*} the neighborhood function h is applied to determine the number of visiting individuals from each other subpopulation. Because of $h_{i^*}(t,k) \in (0,1]$ we can interpret $h_{i^*}(t,k)$ as a *probability for migration* of an individual of the subpopulation Π_k into the actual evaluated Π_{i^*}. The value $r_{i^*,k}$ is defined as the rank of neighborhood $r_{i^*,k} = \text{rank}(\Pi_{i^*}, \Pi_k; \Omega)$ with respect to the topological order Ω between the actually evaluated subpopulation Π_{i^*} and another Π_k. As shown in [41] this approach drastically improves the adaptation rate for EAs.

For selection of the offspring generation we have used a mixture of the well-known (μ,λ)– and the $(\mu+\lambda)$–strategy (in the notation of SCHWEFEL, [35]). While in the (μ,λ)–strategy where μ individuals produce λ children with $\mu < \lambda$ only the μ best of the λ children form the new population, in the $(\mu+\lambda)$–strategy all $\mu+\lambda$ individuals are allowed for the selection process. Whereas in the second strategy the best solution is preserved but the evolution tends to stagnate into a local minimum, in the first one the convergence is decelerated to allow reaching deeper minima (near the global minimum) but good solutions may be lost during the evolution. Balancing the advantages of both strategies [29] in the $(\mu * \lambda)$–approach again μ individuals produce the λ preliminary offsprings [17]. However, in the selection step the μ_t best individuals of the old generation and the λ new ones are allowed for comparison with respect to their fitness to generate the final offspring generation of μ individuals. Thereby μ_t depends on time t of evolution appearing as the number of generations performed:

$$\mu_t = \text{int}\left[(\mu - \mu_\tau)\cdot\gamma(t)\right] + \mu_\tau \quad , \tag{8}$$

with $\text{int}[x]$ stands for the integer value of x. The function $\gamma(t)$ is of decreasing sigmoid type with $0 \le \gamma(t) \le 1$ here chosen as the Fermi function $\gamma(t) = \frac{1}{1+\exp\left(\frac{(t-t_a)}{t_b}\right)}$ to switch near the t_a-th generation from the $(\mu+\lambda)$–strategy to the (μ,λ)–strategy in a definite range of generation steps ($\approx 4t_b$). We have $\mu_0 = \mu$ for the initial value and $\lim_{t\to\infty} \mu_t = \mu_\tau$ ($\mu_\tau \ll \mu$) coding a minimal survival probability for the parent individuals. In this way we get a smoothed switch from the $(\mu+\lambda)$– to the (μ,λ)–strategy, what we call $(\mu * \lambda)$–strategy, combining the advantages of both strategies and, additionally, always preserving the best μ_τ individuals (slightly different from the original (μ,λ)–strategy).

At first we computed $\mathcal{F}$ according to (5) for the original clusters $\mathcal{C}^w$, $\mathcal{C}^{ro}$ and $\mathcal{C}^{rs}$ of the respective standard categories as defined in [4]. Additionally, we computed the weighted concordance coefficient $\tilde{\kappa}$ [1] for comparison with other investigations in the area of psychotherapy research (see Tab. 1). The $\tilde{\kappa}$–coefficients may be interpreted according to Tab. 2 [34]. Applying $\mathcal{F}$ from (5) as fitness measure in the EA we result cluster solutions with higher number of clusters as depicted in Tab. (1): However, the increased $\tilde{\kappa}$-values as well as the better (decreased) $\mathcal{F}$-values refer to a better intra-cluster agreement which correspond to psychological considerations: From a psychological point of view the new clusters show a better coherence which allows a more clear interpretation. Moreover, the increased cluster number allows a more detailed description [1, 25].

From technical point of view we used in our computations $\mu_{\text{all}} = 400$ individuals which were evenly distributed onto 10 subpopulations. The subpop-

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 165-176

database	$\tilde{\kappa}$ original clusters	$\mathcal{F}$ original clusters	$\tilde{\kappa}$ EA clusters	$\mathcal{F}$ EA clusters	$c_{\max}$ EA clusters
P^w	0.334	6.88	0.435	6.24	10
P^{ro}	0.323	6.28	0.421	6.16	9
P^{rs}	0.479	6.16	0.504	6.09	9

Table 1: Values for the weighted concordance coefficient $\tilde{\kappa}$ and the fitness function $\mathcal{F}$ according to the original clusters and the EA generated solutions.

$\tilde{\kappa}$–coefficient	meaning
$\tilde{\kappa} < 0.1$	no agreement
$0.1 \leq \tilde{\kappa} < 0.4$	weak agreement
$0.4 \leq \tilde{\kappa} < 0.6$	clear agreement
$0.6 \leq \tilde{\kappa} < 0.8$	strong agreement
$0.8 \leq \tilde{\kappa}$	nearly complete agreement.

Table 2: Different values for the weighted concordance coefficient $\tilde{\kappa}$ and the respective meaning for intra cluster agreements of the considered observables

ulations were arranged on a ring as topological structure, whereby μ_τ in (8) was given as $\mu_\tau = 1$. We trained the ensemble during $t_{\max} = 5000$ time steps. The characteristic time scale for decreasing the neighborhood between the subpopulation was defined as linear shrinking of $\sigma_h(t)$ in (7) with $\sigma_h(0) = 3$ and $\sigma_h(t_{\max}) = 0.2$.

The choice of the new fitness function $\mathcal{F}$ gives the ability to compare cluster solutions with different cluster number as well as optimization of the cluster number during the evolutionary process. Moreover, the SOM-like migration scheme generated in neighbored subpopulations cluster solutions which are judged by psychotherapist as similar. However, a mathematical proof of this fact is difficult because of the complicate structure of the fitness function but should subject of further considerations as well as an improvement of the fitness function.

References

[1] C. Albani, T. Villmann, B. Villmann, A. Körner, M. Geyer, D. Pokorny, G. Blaser, and H. Kächele. Kritik und erste Reformulierung der kategorialen Strukturen der Methode des Zentralen Beziehungs–Konflikt–Themas (ZBKT). *Psychotherapie, Psychosomatik und Medizinische Psychologie*, 49(11):408–421, 1999.

[2] S.-I. Amari. Topographic organization of nerve fieldsneural fields. *Bulletin of Mathematical Biology*, 42:339–364, 1980.

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 165-176

[3] T. Bäck. *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, New York - Oxford, 1996.

[4] J. Barber, P. Crits-Christoph, and L. Luborsky. A guide to the CCRT Standard Categories and their classification. In L. Luborsky and P. Crits-Chrostoph, editors, *Understanding Transference*, pages 37–50. Basic Books New York, 1990.

[5] H. U. Bauer. Development of oriented ocular dominance bands as a consequence of areal geometry. *Neural Computation*, 7(1):36–50, Jan 1995.

[6] H.-U. Bauer, R. Der, and M. Herrmann. Controlling the magnification factor of self–organizing feature maps. *Neural Computation*, 8(4):757–771, 1996.

[7] H.-U. Bauer, M. Herrmann, and T. Villmann. Neural maps and topographic vector quantization. *Neural Networks*, 12(4–5):659–676, 1999.

[8] H.-U. Bauer and K. R. Pawelzik. Quantifying the neighborhood preservation of Self-Organizing Feature Maps. *IEEE Trans. on Neural Networks*, 3(4):570–579, 1992.

[9] H.-U. Bauer and T. Villmann. Growing a Hypercubical Output Space in a Self–Organizing Feature Map. *IEEE Transactions on Neural Networks*, 8(2):218–226, 1997.

[10] R. Brause and F. Friedrich. A neuro-fuzzy approach as medical diagnostic interface. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'2000)*, page in this volume, Brussels, Belgium, 2000. D facto publications.

[11] J. Buhmann and H. Kühnel. Vector quantization with complexity costs. *IEEE Transactions on Information Theory*, 39:1133–1145, 1993.

[12] P. Crits-Christoph and A. Demorest. List of standard categories (edition 2). In L. Luborsky and H. Kächele, editors, *Der Zentrale Beziehungskonflikt - ein Arbeitsbuch*. PSZ Verlag, Ulm, Germany, 1988.

[13] F. Ermini and C. Marchesi. Intelligent data retrieval from multidimensional clinical archives. In E. Ifeachor, A. Sperduti, and A. Starita, editors, *Neural Networks and Expert Systems in Medicine and Healthcare*, pages 295–303. World Scientific, Singapore, New Jersey, London, Hong Kong, 1998.

[14] D. B. Fogel. *Evolutionary Computation: Towards a New Philosophy of Machine Intelligence*. IEEE Press, Piscataway, NJ, 1995.

[15] T. Greapel and K. Obermayer. A stochastic self-organizing map for proximity data. *Neural Computation*, 11(1):139–155, 1999.

[16] S. Haykin. *Neural Networks - A Comprehensive Foundation*. IEEE Press, New York, 1994.

[17] K. Hering, R. Haupt, and T. Villmann. Hierarchical Strategy of Model Partitioning for VLSI–Design Using an Improved Mixture of Experts Approach. In *Proc. Of the Conference on Parallel and Distributed Simulation (PADS'96)*, pages 106–113. IEEE Computer Society Press, Los Alamitos, 1996.

[18] T. Hofmann and J. Buhmann. Pairwise data clustering by deterministic annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(1):1–14, 1997.

[19] G.-J. Hogrefe. *Testkatalog*. Hogrefe-Verlag, Göttingen, 1996/97.

[20] L. Kaufmann and P. Rousseuw. *Finding Groups in Data - A Introduction to Cluster Analysis*. John Wiley Sons, 1990.

[21] J. Kohlmorgen, K.-R. Müller, and K. Pawelzik. Improving Short-Term Prediction with Competing Experts. In *Proc. ICANN'95*, Paris 1995, 1995.

[22] T. Kohonen. Self-organizing formation of topologically correct feature maps. *Biol. Cyb.*, 43(1):59–69, 1982.

[23] T. Kohonen. *Self-Organizing Maps*. Springer, Berlin, Heidelberg, 1995. (Second Extended Edition 1997).

[24] T. Kohonen. Comparison of SOM point densities based on different criteria. *Neural Computation*, 11(8):212–234, 1999.

[25] A. Körner. *Kategorisierung Von Bezeihungsschemata Mi der Methode "Das Zentrale Beziehungskonfliktthema"*. PhD thesis, University Leipzig, Germany, 2000.

[26] W. Liebert. *Chaos und Herzdynamik*. Verlag Harri Deutsch, Frankfurt/M., Germany, 1991.

[27] R. Linsker. How to generate maps by maximizing the mutual information between input and output signals. *Neural Computation*, 1:402–411, 1989.

[28] L. Luborsky. The core conflictual relationship scheme. In N. Freedman and S. Grand, editors, *Communicative Structure and Psychic Structures*. Plenum Press New York, 1977.

[29] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer–Verlag Berlin Heidelberg New York, third, revised and extended edition, 1996.

[30] H. Mühlenbein, M. Gorges-Schleuter, and O. Krämer. Evolution Algorithm in Combinatorial Optimization. *Parallel Computing*, (7):65–88, 1988.

[31] K. Obermayer, G. G. Blasdel, and K. Schulten. A neural network model for the formation and for the spatial structure of retinotopic maps, orientation-and ocular dominance columns. In T. Kohonen, K. Mäkisara, O. Simula, and J. Kangas, editors, *Artificial Neural Networks*, pages 505–511, Amsterdam, Netherlands, 1991. Elsevier.

[32] H. Ritter, T. Martinetz, and K. Schulten. *Neural Computation and Self-Organizing Maps: An Introduction*. Addison-Wesley, Reading, MA, 1992.

[33] H. Ritter and K. Schulten. On the stationary state of Kohonen's self-organizing sensory mapping. *Biol. Cyb.*, 54:99–106, 1986.

[34] L. Sachs. *Angewandte Statistik*. Springer Verlag, 7-th edition, 1992.

[35] H.-P. Schwefel. *Numerical Optimization of Computer Models*. Wiley and Sons, 1981.

[36] T. Villmann. Application of evolutionary algorithms for clustering of non-metric data in medicine. In G. Brewka, R. Der, S. Gottwald, and A. Schierwagen, editors, *Fuzzy-Neuro-Systems'99 (FNS'99), Proc. Of the FNS-Workshop*, pages 155–165, Leipzig, 1999. Leipziger Universitätsverlag.

[37] T. Villmann. Topology preservation in self-organizing maps. In E. Oja and S. Kaski, editors, *Kohonen Maps*, number ISBN 951-22-3589-7, pages 279–292, Amsterdam (Holland), June 1999. Helsinki, Elsevier.

[38] T. Villmann. Modelling of the cortical deformations in the auditive cortex for tinnitus disease using self-orgainzing maps. Technical Report in prep., University Leipzig, Clinic for Psychotherapy, Leipzig, Germany, 2000.

[39] T. Villmann and H.-U. Bauer. Applications of the growing self-organizing map. *Neurocomputing*, 21(1-3):91–100, 1998.

[40] T. Villmann, R. Der, M. Herrmann, and T. Martinetz. Topology Preservation in Self–Organizing Feature Maps: Exact Definition and Measurement. *IEEE Transactions on Neural Networks*, 8(2):256–266, 1997.

[41] T. Villmann, R. Haupt, K. Hering, and H. Schulze. Parallel evolutionary algorithms with som-like migration. In A. Dobnikar, N. Steele, D. W. Pearson, and R. Albrecht, editors, *Artificial Neural Networks and Genetic Algorithms (Proc. Of ICANNGA'99)*, pages 274–279, Wien - New York, 1999. Springer-Verlag.

[42] T. Villmann and M. Herrmann. Magnification control in neural maps. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'98)*, pages 191–196, Brussels, Belgium, 1998. D facto publications.

[43] G. Westhoff. *Handbuch Psychosozialer Meßinstrumente*. Hogrefe-Verlag, Göttingen, 1993.

[44] D. J. Willshaw and C. V. der Malsburg. How patterned neural connections can be set up by self–organization. *Proceedings of the Royal Society of London, Series B*, 194:431–445, 1976.

[45] A. Wismüller, F. Vietze, D. Dersch, K. Hahn, and H. Ritter. A neural network approach to adaptive pattern analysis - the deformable feature map. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'2000)*, page in this volume, Brussels, Belgium, 2000. D facto publications.