

Parallel Evolutionary Algorithms with SOM-Like Migration and their Application to Real World Data Sets

Th. Villmann[◊], R. Haupt[•], K. Hering[•] and H. Schulze[•]

Universität Leipzig, Germany

[◊]Klinik für Psychotherapie, 04107 Leipzig, K.-Tauchnitz-Str.25
villmann@informatik.uni-leipzig.de

[•]Institut für Informatik, 04109 Leipzig, Augustusplatz 10/11
{haupt,hering,schulze}@informatik.uni-leipzig.de

Abstract

We introduce a multiple subpopulation approach for parallel evolutionary algorithms the migration scheme of which follows a SOM-like dynamics. We successfully apply this approach to clustering in both VLSI-design and psychotherapy research. The advantages of the approach are shown which consist in a reduced communication overhead between the subpopulations preserving a non-vanishing information flow.

1 Introduction

Evolutionary Algorithms (EAs) are a biologically motivated stochastic iterative optimization method. In EAs the manipulation of objects, which are called individuals $s \in \mathbf{\Pi}$, is separated from their evaluation by a fitness 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.¹ If in one time step all individuals have gone under manipulation and evaluation, from these the *selection operator* extracts a new generation for the next iteration step. For a detailed overview we refer to [10].

Several approaches were developed to improve this basic EA-scheme. Especially, multiple subpopulation approaches are widely considered [4, 11]. Thereby the basic population $\mathbf{\Pi}$ is divided into subpopulations $\mathbf{\Pi}_i$

which have more or less communication during the evolution which may be realized as migration [15].

Yet, high communication frequency causes large effort, especially if one uses a multi-processor system for the evaluation in such a way that each subpopulation is evaluated by one processor. However, the high information flow through the set $\mathbf{\Pi}$ of subpopulations increases the genetic diversity which, in general, accelerates the development of the fitness of the individuals during the first generations. In contradiction, in the convergence phase the problem of homogeneity between the subpopulations may occur. This problem is more relevant for discrete optimization tasks, however it is also addressed elsewhere [4, 10, 11]. On the other hand, if we allow only a small amount of communication the performance of a multi-processor system will be improved. Moreover, because of the relative independence between the subpopulations each of them searches in a different region of the solution space. However, the improvement rate of the fitness of the individuals during the first generation is reduced which leads, in general, to a lower fitness.

Therefore, in the present contribution we focus on a merging of both strategies. For this purpose we apply the concept of collective learning with a progressing separation during the time development as it is known from the concept of *neural maps* considered by KOHONEN [8]. In addition, we prefer a mixture of both the (μ, λ) - and the $(\mu + \lambda)$ -strategy (in the notation of SCHWEFEL, [13]) for the selection operator. We show the success of this approach for two real world applications (in the area of VLSI-design and psychotherapy research).

¹Thereby, the fitness function may contain explicit expert knowledge regarding to the optimization task which may be difficult to code otherwise.

2 EAs with SOM-Like Migration for Clustering

2.1 Clustering as a Partitioning Task

Mapping a set \mathcal{S} onto clusters $C \in \mathcal{C}$ in the context of EAs can be taken as a *partitioning problem* under constraints which are specified by the fitness function F . Thereby, a *partitioning* of a nonempty set \mathcal{S} related to a nonempty set \mathcal{C} is a unique and surjective mapping

$$\Phi : \mathcal{S} \rightarrow \mathcal{C} \quad (2.1)$$

Then a *partition* Ψ_Φ of \mathcal{S} related to the partitioning Φ is given by

$$\Psi_\Phi = \{\Phi^{-1}(C) \mid C \in \text{cod}(\Phi)\} \quad (2.2)$$

whereby $\text{cod}(\Phi)$ is the range of Φ [7].

For solving the partitioning task an individual in a generation of an EA describes a certain partition. In the present paper we assume that the number of clusters to be built is predefined as c_{\max} . Furthermore, we consider \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 a cut of two individuals at the same point followed by crossed sticking together.

2.2 EAs with SOM-Like Migration Scheme

We consider a *multiple subpopulation approach* whereby a set $\mathbf{\Pi}$ of subpopulations Π_i is arranged on a topological order which is often chosen to be a regular lattice, for instance a ring or a quadratic lattice². Between these subpopulations a *migration scheme* was introduced, originally developed by TOTH&LÖRINCZ [15] in a basic variant, and applied here in an extended approach. In this extended approach a visit (migration) from individuals between neighboring subpopulations is allowed for a short time *regarding to the topological order* Ω in $\mathbf{\Pi}$ with respect to a time-dependent neighborhood function

$$h_{i^*}(t, k) = (1.0 - \epsilon_h) \cdot \exp\left(-\frac{r_{i^*,k}}{2(\sigma_h(t))^2}\right) + \epsilon_h \quad (2.3)$$

with a small positive number ϵ_h . During the evalua-

²In general, other arrangements are also admissible. Then the lattice can be defined by a connection matrix which describes the neighborhood relations.

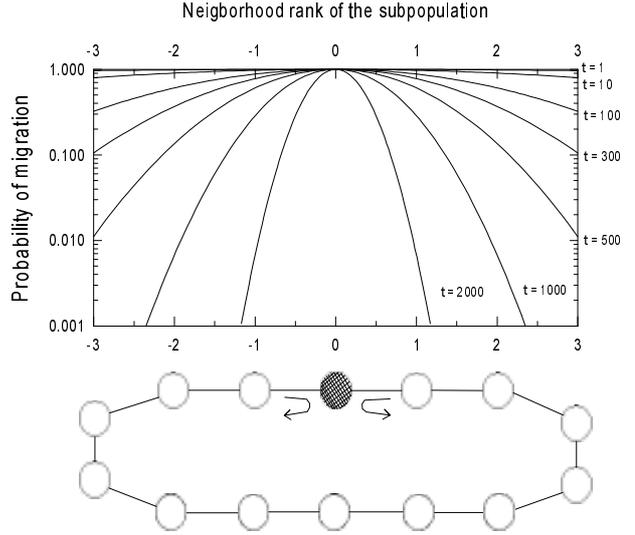


Fig. 1: Plot of the neighborhood function $h_{i^*}(t, k)$ from eq.(2.3) determining the migration scheme.

tion 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

$$r_{i^*,k} = \text{rank}(\Pi_{i^*}, \Pi_k, \Omega) \quad (2.4)$$

between the actually evaluated subpopulation Π_{i^*} and another Π_k according to the topological order Ω in $\mathbf{\Pi}$. At the beginning t_0 of the evolution process the range h of the neighborhood nearly comprises the complete set of subpopulations and decreases exponentially during the time determined by

$$\sigma_h(t) \xrightarrow{t \rightarrow \infty} 0 \quad (2.5)$$

in (2.3) (see Fig. 1).

Following this approach one can take the individuals of all subpopulations at the beginning as an uniform population which performs a first rough adaptation process. During the further development the various subpopulations become more and more separated from each other and, hence, search in *different regions* of the solution space. This approach of first rough adaptation of neighboring subpopulations together with a more fine tuning in the further process by more but not completely independent subpopulations arranged on a topological structure shows

an analogy to the concept of *self-organizing maps* (SOM) in the field of neural maps [8]. Thereby, the neural units are also placed on a topological structure (lattice) and the collective dynamics during the learning procedure follows the same idea as the subpopulations in the present article: a first rough adaptation of the neuron weights takes place changing to a precise adjustment simultaneously with loosing the strong neighborhood conditions in the lattice.

On the other hand, ϵ_h in (2.3) preserves a remaining probability for migration. In this way one has a non-vanishing information flow through the topological ordered set $\mathbf{\Pi}$ of subpopulations which *accelerates* the adaptation process, especially, if the search space possesses many local minima [10] (see sect. 4.2). Thereby, neighboring subpopulations contain similar in fitness but genetically different individuals whereas usual EAs tend to homogeneity. Hence, if the fitness function F only incompletely describes the expert knowledge as in sect. 4.1, the expert can choose one of these similar individuals as the final solution. In the present applications the topological order in $\mathbf{\Pi}$ for the SOM-like migration scheme is a ring.

3 The $(\mu * \lambda)$ -Approach for the Selection Scheme in EAs

For selection of the offspring generation we have used a mixture of the (μ, λ) - and the $(\mu + \lambda)$ -strategy (in the notation of SCHWEFEL, [13]). 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 [10] the $(\mu * \lambda)$ -approach was introduced to solve hard partitioning problems in VLSI-design [7] with many widely distributed local minima.

In the $(\mu * \lambda)$ -approach again μ individuals produce the λ preliminary offsprings. 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:

$$\mu_t = \text{int}[(\mu - \mu_\tau) \cdot \gamma(t)] + \mu_\tau \quad , \quad (3.1)$$

with $\text{int}[x]$ stands for the integer value of x . The function $\gamma(t)$ is of decreasing sigmoid type with $0 \leq \gamma(t) \leq 1$ here chosen as the Fermi function

$$\gamma(t) = 1 / (1 + \exp((t - t_a)/t_b)) \quad (3.2)$$

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

$$\lim_{t \rightarrow \infty} \mu_t = \mu_\tau \quad \text{with} \quad (\mu_\tau \ll \mu) \quad (3.3)$$

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).

4 Applications and Results

4.1 Clustering of Psychological Categories

One of the mostly used methods for 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 [9]. The method investigates so-called *relationship-episodes*, which are often reported by the patients in their therapeutical sessions. For each of these episodes the components *wish of the subject* (W), *response of the object* (RO) and *response of the subject* (RS) are encoded which then are used to perform the CCRT. BARBER ET AL. [2] determined a system \mathcal{S}^w of $s_{\max}^w = 34$ so-called *standard categories* S_j^w to classify the wishes which are collected in a set \mathcal{C}^w of $c_{\max}^w = 8$ clusters C_k^w [2]. The number and the interpretation of the clusters as well as the assignment of the standard categories are obtained from the experience of several psychotherapists using conventional statistical methods. Analogously they explained $s_{\max}^{ro} = 30$ categories $S_j^{ro} \in \mathcal{S}^{ro}$ for encoding the RO and $s_{\max}^{rs} = 30$ categories $S_j^{rs} \in \mathcal{S}^{rs}$ for encoding the RS which are collected in $c_{\max}^{ro} = 8$ clusters $C_k^{ro} \in \mathcal{C}^{ro}$ and $c_{\max}^{rs} = 8$ clusters $C_k^{rs} \in \mathcal{C}^{rs}$, respectively.

However, as mentioned in [1] the clusters are still correlated again what *leads to low reliability rates*.

Hence, the task is to reform the clusters of standard categories to improve the reliability. For this purpose several raters judge a large number of episodes and determine the most relevant standard category $S_{j^*}^w(i)$ of each wish and, *in addition*, a second one (denoted as $S_{j^+}^w(i)$), which has to be different from the first one but is also well describing. All pairs $p_i^w = (S_{j^*}^w(i), S_{j^+}^w(i))$, $i = 1 \dots N_w$, $N_w = 7922$, form a database P^w which implicitly contains the correlation information between the standard categories. Analogously we obtained a database P^{ro} of $N_{ro} = 9651$ pairs p_i^{ro} and a database P^{rs} of $N_{rs} = 11629$ pairs p_i^{rs} .

If the respective clusters are determined in a reliable way according to an arbitrary clustering algorithm, both the most and the second relevant standard category should belong to the same cluster after this procedure, otherwise we denote this fact as misclassification. Let us denote

$$\phi_{j^*}^i = \Phi(S_{j^*}^w(i)) \quad , \quad \phi_{j^+}^i = \Phi(S_{j^+}^w(i)) \quad (4.1)$$

for a given cluster solution according to (2.1). Instead of applying the misclassifications \mathcal{M} , we involved the conditional probabilities

$$\rho_{j^*,j^+}^i = p(S_{j^*}^w(i) | S_{j^+}^w(i)) \quad , \quad (4.2)$$

estimated from the database, into the fitness function:

$$\hat{F}(s) = \sum_{\substack{i=1 \\ \phi_{j^*}^i \neq \phi_{j^+}^i}}^N \max(\rho_{j^*,j^+}^i, \rho_{j^+,j^*}^i) \quad (4.3)$$

However, we have to pay attention to the balancing of the clusters which is measured by the variance of the number of the standard categories belonging to the clusters as suggested in [16]. Without this constraint the optimal solution would be to collect all standard categories into only one cluster. Furthermore, some explicitly known therapeutical knowledge about the parallel appearance of standard categories is coded. Both aspects are included in an additional term \tilde{F} . Thus the final fitness function is obtained by

$$F = \hat{F} + \tilde{F} \quad . \quad (4.4)$$

The cluster solution found by our approach shows more coherence than the original one from a psychological point of view. The clusters itself are more consistent and, on the other hand, now there is a greater differentiation between the clusters in psychological meaning. For a more detailed psychological consideration we refer to [1].

Table 1: Different values for the weighted concordance coefficient κ and the respective meaning for the agreement of the appearance of the considered observables

κ -coefficient	meaning
$\kappa < 0.1$	no agreement
$0.1 \leq \kappa < 0.4$	weak agreement
$0.4 \leq \kappa < 0.6$	clear agreement
$0.6 \leq \kappa < 0.8$	strong agreement
$0.8 \leq \kappa$	nearly complete agreement.

In addition, the new clusters are compared with the original ones via the *weighted* concordance coefficient κ as a mathematical measure [3] which scores the simultaneous appearance of the respective clusters $\phi_{j^*}^i$ and $\phi_{j^+}^i$ of a pair $p_i \in P$ obtained according to (4.1). At first we computed κ for the original clusters \mathcal{C}^w , \mathcal{C}^{ro} and \mathcal{C}^{rs} of the respective standard categories as defined in [2]. We found the values $\kappa_w = 0.334$, $\kappa_{ro} = 0.330$ and $\kappa_{rs} = 0.492$ whereby the standard deviations were $\nu_w = 0.00626$, $\nu_{ro} = 0.00600$ and $\nu_{rs} = 0.00601$, respectively. Hence, the κ -values for W and RO are related to an only weak agreement whereas the κ_{rs} refers to a clear agreement according to Tab.1 [12]. The values κ_w , κ_{ro} and κ_s correspond to $\mathcal{M}_w = 4662$ (58.8%), $\mathcal{M}_{ro} = 5410$ (56.1%) and $\mathcal{M}_{rs} = 4910$ (46%) misclassifications in the respective databases P^w , P^{ro} and P^{rs} . Our parallel SOM-like EA-approach together with the $(\mu * \lambda)$ -strategy yields the following results:

	\mathcal{M}	κ	ν
P^w	3764 (47.5%)	0.429	0.00675
P^{ro}	4917 (50.9%)	0.421	0.00578
P^{rs}	4352 (37.4%)	0.562	0.00525

(4.5)

In fact, now all κ -values refer to a clear agreement according to Tab.1 which is a strong improvement. Here we mention again that we did not use κ as fitness function itself because of the more adequate description (in psychotherapeutic sense) by the conditional probabilities. However, κ is the mostly used mathematical measure in psychotherapy research. Hence we have to present it for comparison.

4.2 VLSI Model Partitioning for Logic Simulation

For the logic design of whole microprocessor structures time-extensive cycle-based simulation processes

are necessary [14]. Time spent for simulation can be drastically reduced using *parallel simulation* based on *model partitioning*. The model partitioning problem can be formulated as a combinational optimization problem. In this context partitions are characterized by a complex cost function [5] which estimates the run-time of one parallel simulation cycle of the corresponding hardware model parts. This cost function has to be minimized to reduce the expected parallel simulation time. Hierarchical model partitioning for parallel system simulation allows a successful application of EAs at the second hierarchy level [7].

The components of a suited prepartition (super-cones) obtained from the first hierarchy level are used as basic set for the partitioning at the second level yielding a final set of clusters (blocks). To apply EAs at the second level, a set of initial partitions is required as start population. They are built by special algorithms, for instance by *MOCC* [7], using expert knowledge for qualified but genetically different individuals (i.e. partitions) to start with a good population of high diversity. Each component of an individual (partition) represents a super-cone coded by an integer as block index. The *fitness* F of an individual is identified by the *estimated run-time* for parallel simulation with respect to the corresponding partition and is calculated using two kinds of hypergraphs – *Overlap and Communication Hypergraph* as explained in [6].

To achieve better partitions in shorter time we have parallelized EAs using the multiple subpopulation approach with the SOM-like migration scheme mentioned above. For each subpopulation the $(\mu * \lambda)$ -strategy is applied. The migration is implemented by nonblocking Point-to-Point communication in the frame of the *Message Passing Interface*.

Experimental results are given for an *IBM S/390* processor model which is partitioned into 15 blocks via the *STEP* prepartitioning algorithm resulting in 250 super-cones [5]. At the second hierarchy level three different EA strategies are realized with the parameters $\mu = 294$, $\lambda = 700$ in each run. A sequential one-population EA is opposed to two parallel multiple subpopulation EAs each consisting of 7 subpopulations (i.e. $\tilde{\mu} = 42$, $\tilde{\lambda} = 100$ for each subpopulation). We compare the SOM-like migration scheme with an *all-to-all communication scheme* where the communication effort is much higher than for the SOM-like case. Considering the partitioning effort for the multiple subpopulation approach, in Fig.2 the fitness of the best individual is plotted over the time t_{part} spent for the EA partitioning.

Parallel partitioning drastically reduces the parti-

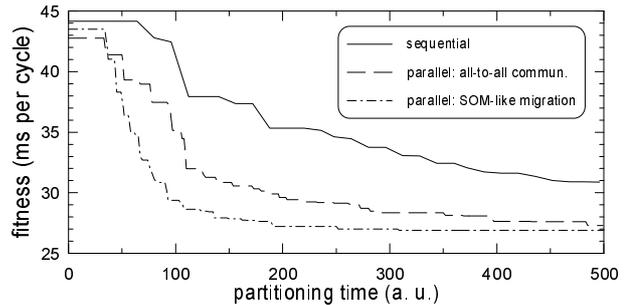


Fig. 2: Fitness of the best individual of all subpopulations comparing a sequential run with two parallel ones in dependence on the *partitioning time* t_{part} measured in arbitrary units (a. u.).

tioning time because the partitioning effort is distributed to 7 processors. The fitness of the best individual significantly faster decreases than in the sequential case. But the all-to-all communication scheme is accompanied by a high communication overhead. Using our SOM-like migration approach this communication effort can be reduced in such a way that better individuals (partitions) are obtained in shorter partitioning time t_{part} . So, in the example discussed here an estimated run-time of 27 ms per cycle (fitness) is reached in the half of the partitioning time comparing to the all-to-all communication scheme. Although the individuals of our initial population are already equipped with expert knowledge and not randomly produced, EAs yield a reduction of estimated run-time from ≈ 44 ms down to 27 ms (see Fig.2).

5 Concluding Remarks

We have developed a SOM-like migration scheme for multiple subpopulation systems for EAs which is inspired by neuron dynamics of self-organizing maps in the area of neural computation. Thereby the subpopulations are arranged on a topological structure and the possibility for migration depends on the neighborhood rank and decreases during time with a remaining rest probability. Additionally, we have introduced a balance between the classical $(\mu + \lambda)$ - and (μ, λ) -strategy. We successfully applied this approach to two real world data sets from VLSI-design and psychotherapy research and demonstrated its advantages.

Acknowledgments. The authors would like to thank D. Pokorny (Universität Ulm) for giving the approach of the weighted concordance coefficient and C. Albani / A. Körner (Universität Leipzig) for psychological data and discussions. Moreover, R. Haupt and H. Schulze are grateful for the support by DEUTSCHE FORSCHUNGSGEMEINSCHAFT (DFG) under grant SP 487/1-2.

References

- [1] C. Albani, B. Villmann, T. 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 - G.-Thieme-Verlag Stuttgart, New York*, page to appear, 1998.
- [2] J. Barber, P. Crits-Christoph, and L. Luborsky. A guide to the CCRT Standard Categories and their classification. In L. Luborsky and P. Crits-Christoph, editors, *Understanding Transference*, pages 37–50. Basic Books New York, 1990.
- [3] J. Cohen. Weighted kappa. *Psychological Bulletin*, 70:213–220, 1968.
- [4] D. B. Fogel. *Evolutionary Computation: Towards a New Philosophy of Machine Intelligence*. IEEE Press, Piscataway, NJ, 1995.
- [5] R. Haupt, K. Hering, U. Petri, and T. Villmann. Hierarchical model partitioning for parallel VLSI-simulation using evolutionary algorithms improved by superpositions of partitions. In K. Lieven, editor, *Proceedings of European Congress on Intelligent Techniques and Soft Computing (EUFIT'97)*, pages 804–808, Aachen, Germany, 1997.
- [6] R. Haupt, K. Hering, and T. Siedschlag. Integration of a Local Search Operator into Evolutionary Algorithms for VLSI-Model Partitioning. In K. Lieven, editor, *Proceedings of European Congress on Intelligent Techniques and Soft Computing (EUFIT'98)*, pages 377 – 381, Aachen, Germany, 1998. ELITE Foundation.
- [7] 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.
- [8] T. Kohonen. *Self-Organizing Maps*. Springer, Berlin, Heidelberg, 1995. (Second Extended Edition 1997).
- [9] L. Luborsky. The core conflictual relationship scheme. In N. Freedman and S. Grand, editors, *Communicative Structure and Psychic Structures*. Plenum Press New York, 1977.
- [10] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag Berlin Heidelberg New York, third, revised and extended edition, 1996.
- [11] H. Mühlenbein, M. Gorges-Schleuter, and O. Krämer. Evolution Algorithm in Combinatorial Optimization. *Parallel Computing*, (7):65–88, 1988.
- [12] L. Sachs. *Angewandte Statistik*. Springer Verlag, 7-th edition, 1992.
- [13] H.-P. Schwefel. *Numerical Optimization of Computer Models*. Wiley and Sons, 1981.
- [14] W. G. Spruth. *The Design of a Microprocessor*. Springer Berlin, Heidelberg, 1989.
- [15] G. J. Tóth and A. Lőrincz. Genetic algorithm with migration on topology conserving maps. In S. Gielen and B. Kappen, editors, *Proc. ICANN'93, Int. Conf. on Artificial Neural Networks*, pages 605–608, London, UK, 1993. Springer.
- [16] T. Villmann, B. Villmann, and C. Albani. Application of evolutionary algorithms to the problem of new clustering of psychological categories using real clinical data sets. In B. Reusch, editor, *Computational Intelligence - Theory and Applications - Proc. Of the International Conference on Computational Intelligence - 5th. Dortmund Fuzzy-Days*, pages 311–320, Berlin, New York, Heidelberg, 1997. Lecture Notes in Computer Science 1226, Springer-Verlag.