ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 215-220

Three Algorithms for Searching the Minimum Distance in Kohonen Maps

Viktor Tryba[1], Karl Goser[2]

1 SICAN GmbH, Garbsener Landstrasse 10, 3000 Hannover 21, on leave from 2
2 Lehrstuhl Bauelemente der Elektrotechnik, Emil-Figge-Strasse 68, Universität Dortmund

Abstract: Three algorithms for searching the minimum distance in self-organizing maps are presented and analysed in regard to global communication of cells and necessity of threshold functions. A new algorithm for searching the minimum distance without threshold function is presented.

1. Introduction

The principle of the self-organizing map was discovered by Kohonen [1] and has been used for many applications (see overview in [2,3]). Fig. 1a shows a one-dimensional self-organizing map which is a regular array of cells that are located at the discrete positions x ($1 <= x <= I$, $x \in \mathcal{N}$). Each cell stores a weight vector $W_x=(w_{x1},w_{x2}, .. , w_{xk}, .. ,w_{xK})$ of K components with $w_{xk} \in \mathcal{R}$. The map learns a set $S=(S_1,S_2, .. ,S_l, .. ,S_L)$ of L input vectors $\in \mathcal{R}^K$. After a random initialization of all weight vectors $W_x(t_0)$ at the beginning t_0 of the training time t_m ($t_0 <= t_m <= t_{end}$), the following steps are executed for each input-vector:

(1) Search the most similar weight-vector to the input-vector.

(2) Adapt the most similar weight-vector and the weight-vectors in a neighbour area around it.

The adaptation area is large at the beginning of the learning and decreases slowly during the training time until only one cell adapts. Various modified algorithms have been published, but in all cases the algorithms consist of these two major steps. The similarity of a weight-vector and the input-vector is often measured by the Euclidian distance. It is then necessary to search the absolute minimum of the distance in the map. In this paper, we will focus on the searching process, because it is a key question for the process of self-organization.

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 215-220

2. Algorithms for searching minimum distance

The basic problem can be seen from Fig. 2 : After the calculation of the distances d_x , the position of the absolute minimum in the distance landscape must be found.

2.1 Minimum detection by direct comparison of all cells

Trying to calculate the result of the learning mathematically meets the problem that there is no explicit formula for the result of the searching process. To write an explicit formula

$$x_{min} = f_{min}(d_x) \tag{1}$$

that calculates the position x_{min} from the known distance values d_x, we will use threshold functions and compare all pairs of cells with each other. Let $\sigma(y)$ be a threshold function, that is 1 for y >= 0, and else 0, and let d_x be the distance of a given cell that we compare with all other distances d_ξ in the map by calculating $\sigma(d_\xi - d_x)$ for all pairs (ξ, x) in the map. The product

$$P_x = \prod_{\xi} s(d_\xi - d_x) \ , \quad \xi \neq x \tag{2}$$

is 1, if d_x is smaller than all d_ξ , else 0. If there is only one absolute minimum, then

$$x_{min} = \sum_{x} x\, P_x \tag{3}$$

indicates the position of the absolute minimum as an explicit formula. It happens occasionally that there are several cells with the same and minimum distance. It is usually suggested to make a random choice of the most similar cell. Simulations show that it is not important how the center of the adaptation function is chosen as long as only one cell is selected. It is possible to choose the average of their coordinates. It may also be argued that the discontinuity of the treshold function $\sigma(y)$ makes the algorithm sensitive to random effects and unpredictable. It is therefore important to state that the threshold function $\sigma(y)$ may be replaced by a smoother threshold function, e.g. the fermi-function. Simulations show that steep threshold functions are only necessary at the end of the training process and that the threshold function can be smooth at the beginning of the learning. The steepness of the function is then gradually increased during the training. With (2) and (3), it becomes clear that the process of searching the absolute minimum needs a large number of threshold functions, and global communication of all the distance values to all other cells. The direct hardware implementation of this mathematical formula would require a large hardware effort, because each operation in a mathematical formula and the transfer of information consumes time and energy in any physical or biological realization.

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 215-220

2.2 PID-Controller for minimum detection

Fig. 3 presents a different approach to find the absolute minimum that is more economic in regard to global communication and the number of threshold functions. It is supposed that D_G is a global distance value that is broadcasted to all cells and that each cell compares its local distance value d_x with D_G according to

$$a_x = \sigma(D_G - d_x) . \qquad (4)$$

All values a_x of all cells are summed in one large adder for the map with $A = \Sigma a_x$. The value of A can be understood as the number of firing cells. The minimum can be detected if D_G is changed such that A becomes 1. This can be achieved by one PI or PID-controller for the map. The time for the minimum detection mainly depends on the quality of the designed controller and is almost independant of the number of cells. The type of information that is exchanged between the cells is quite different from the searching process of 2.1 : The global value D_G that must be broadcasted to all cells, and the sum A of the number of firing cells that must be known to the controller.

2.3 Searching the minimum distance by the time-dependant Schrödinger equation

It seems to be difficult to indicate a search algorithm without visible or hidden threshold functions. A different strategy to find the absolute minimum is the use of networks where the output of the cells are fed back to the input. Threshold function here often have the task to avoid divergence of the output of the winning cells in the network. We will now introduce an algorithm to search the absolute minimum without non-linearity or threshold function that is stable by itself. Also, there is only communication to direct neighbour cells. The basic idea is to look for a process in physics that tends to find the absolute minimum and use it as an analogy for the map. Preliminary experiments with partial differential equations that are derived from the equation for thermal diffusion processes had not been succesful, because a diffusion process converges towards a constant function after a long time. In quantum mechanics, however, there is an interesting process that tends to find the absolute minimum: If an electron of mass m is localized in a local minimum of an electric potential $W_{pot}(x)$ (Fig 4 a), and there is a deeper minimum near by, then a certain probability exists for the electron to tunnel trough the potential wall and find the absolute minimum [4-6]. The behaviour of the electron is probabilistic in principal and thus only a probability $|\psi(x,t)|^2 dx$ to find the electron in an intervall dx around x can be assigned to the electron. Fig. 4 b and c show the probability density before and after the transition. A similar mechanism has already been used by Rujan [7] to solve optimisation problems. He reports excellent results for examples of the travelling salesman problem. The function $\psi(x,t)$ can be calculated from the Schrödinger equation

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 215-220

$$\frac{h}{2m}\left(\frac{d^2\psi}{dx^2}+\right) - (1/h)\, W_{pot}(x)\,\psi = -i\frac{d\psi}{dt} \qquad \xi \in \mathcal{R} \qquad (5)$$

(i imaginary number, h=1.0545 10^{-34} Ws^2 as Planck constant). The complex variable ψ denotes the wave-function and has no explicit physical meaning but is a function that can be used to derive probability distributions of the real physical values of the electron. Usually, one has to normalize ψ so that the total probability is unity. The Schrödinger equation will conserve this probabilty over the time. Depending on ψ, W_{pot}, h and m, the transision from the local minimum to the deeper minimum can be similar to a diffusion process, but it can also have wave-like properties.

To use this principle for searching the minimum in the map, we assume that the distance d_x corresponds to the electric potential $W_{pot}(x)$. At the beginning of the searching process, $\psi(x,t)$ is initialized by an arbitrary constant, and then the Schrödinger equation (5) is solved numerically (refer to [5,6] for numerical techniques). The boundary conditions are $\psi_{-1} = \psi_{+1}$ and $\psi_{I+1} = \psi_{I-1}$ for the cells at x=1 and at x =I. To assure an easy transision through the potential walls, it is useful to start with a large value for h at the beginning at the searching process and decrease h gradually during the searching. This corresponds to a reduction of the energy during the searching process. Also the lateral extension of $|\psi(x,t)|^2$ should contract during the training process (cf. Fig. 1 (b)). Fig. 4 d shows the function which was chosen for h during the searching process of several input-vectors. With this choice, the maximum of $|\psi(x,t)|^2$ usually settles in the deepest and widest minimum. The maximum of $|\psi(x,t)|^2$ indicates the approximate position of x_{min}, but for the adaptation of the map it is not necessary to search this position because $|\psi(x,t)|^2$ is taken directly as the adaption function. The learning algorithm for the map becomes now for each input vector:

1) Calculate the distance d_x for all cells.

2) Initialize $\psi(x,t)$ with an arbitrary constant and solve the time-dependant Schrödinger equation numerically .

3) Adapt the weight vectors proportional to $|\psi(x,t)|^2$.

With this algorithm, self-organization similar to Kohonen maps can be achieved ([8,9]). The interesting feature of this algorithm is that both the searching process and the creation of the adaption function are done by solving a partial differential equation that is known from physics. With this analogy, the Schrödinger equation may be understood as a partial differential equation that describes information transportation in the map. Various interesting new questions arise from this analogy, e. g. if there are any similarities of biological information processing and

quantum mechanics, and if the transport and localisation of information has any similarities with the transportation of an electron that must be considered by quantum mechanics.

3. Conclusion

We have presented three methods of searching the minimum distance in a self-organizing map. Searching the minimum distance is a key question to self-organization which deserves detailed investigation. With the algorithm in 2.3, it is possible to give a searching algorithm without threshold functions and with communication via direct neighbored cells only. The results suggest that the interpretation of the adaptation function as a thermal diffusion process is perhaps not appropiate. Simulated tunneling is a searching process that is different from simulated annealing and deserves more interest for optimization problems. The interpretation of the searching algorithm in chapter 2.3 may lead to fundamental questions of information processing.

4. References

1. Kohonen, T., Automatic Formation of Topological Maps of Patterns in a Self-organizing System, proceedings of the 2nd Scandinavian Conference on Image Analysis, Helsinki, June 15-17, 1981, pp. 1-7.
2. Kohonen, T., The Self-Organizing Map, Proceedings of the IEEE, Vol. 78, Sept. 1990, pp. 1464-1480.
3. Goser, K., Kohonen´s Maps - Their Application and Implementation in Microelectronics Proceedings of ICANN ´91, June 24-28 1991, pp. 703-708.
4. Bransden, B. H., Joachain, C. J., Introduction to Quantum Mechanics, Longman Scientific & Technical, Copublished with John Wiley & Sons, Inc., New York, 1989.
5. Goldberg, A., Schey, H., Schwartz, J., Computer Generated Motion Pictures of One-Dimensional Quantum-Mechanical Transmission and Refection Phenomena, American Journal of Physics, Vol. 3, 1967, pp. 177-186.
6. Galbraith, I., Yin Sing, C., Abraham, E., Two-Dimensional Time-Dependent Quantum-Mechanical Scattering Event, American Journal of Physics, Vol. 52, No. 1, 1984, pp. 177-186.
7. Rujan, P., Searching for optimal configurations by simulated tunneling, Condensed Matter 73, Zeitschrift für Physik B, Springer Verlag 1988, pp. 391-416.
8. Tryba, V., Goser, K., A modified Algorithm for Self-Organizing Maps based on the Schrödinger Equation, Proceedings of IWANN 91, Granada, Spain, Sept. 91, Lecture Notes in Computer Science, Springer Verlag, 1991, pp. 33-47.
9. Tryba, V., Selbstorganisierende Karten, Theorie, Anwendung und mikro-elekronische Realisierung, Dissertation Universität Dortmund, Jul. 1992, VDI-Verlag, Reihe Fortschrittsberichte, 1992.

ESANN'1993 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 7-8-9 April 1993, D-Facto public., ISBN 2-9600049-0-6, pp. 215-220

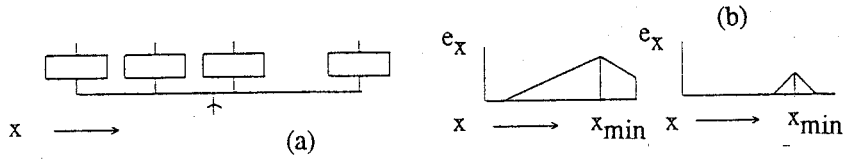

Fig. 1 (a) Self-organizing map with cells that are arranged in one dimension,
(b) adaptation function during the training of the map.

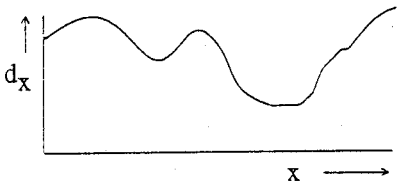

Fig. 2 Distance landscape with absolute minimum

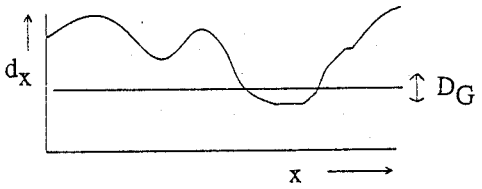

Fig 3. Minimum detection by a distance controller

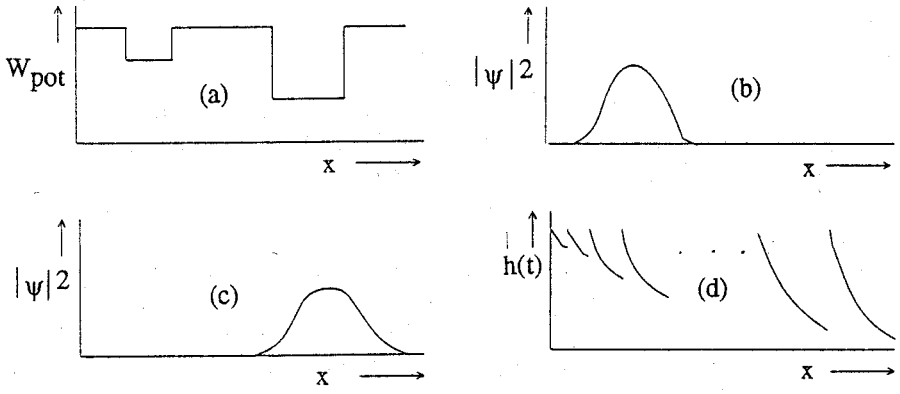

Fig 4. (a) Electric potential with a local and an absolute minimum , (b) probability density for an electron in the local minimum and (c) in the absolute minimum, (d) function h(t) that was used instead of h during the training of the map.