Self-Organizing Recognition and Classification of Relational Structures

Brijnesh J. Jain (bjj@cs.tu-berlin.de)
Department of Computer Science; Technical University Berlin
Germany

Fritz Wysotzki (wysotzki@cs.tu-berlin.de)
Department of Computer Science; Technical University Berlin
Germany

Abstract

We present a novel self-organizing structure recognition (SOSR) network for classification and recognition of relational structures represented by graphs. The system consists of several subnets each comparing an input structure with a given model structure. The subnets are indirectly coupled via a winner-take-all (WTA) classifier. During classification the SOSR system deactivates subnets which indicate large dissimilarities between the input structure and the corresponding models. First experiments show that this mechanism significantly reduces the computational effort in comparison to traditional classification systems using a comparative maximum selector as a classifier.

Introduction

We describe a hierarchical neural net for the recognition and classification of relational structures by matching with class prototypes which was primarily developed from a theoretical point of view and for practical applications in Artificial Intelligence. Classification by means of prototypes is well known in the psychological literature (e.g. Rosch, 1975; Rosch and Lloyd, 1978) but usually is modeled using feature vectors as description of objects and prototypes, respectively. In the context of modeling semantic memory and discussion of the binding problem in Cognitive Neuroscience relational descriptions and representations of structured objects play nowadays a major role (e.g. Hinton, 1994; Taylor, 1993; Taylor, 1996; Singer, 2000). Seen from the point of view of modeling the dynamics of neural structures in connection with psychologically observed behavior we are not primarily interested in the neural (population or assembly) code of representing relations (e.g. Singer, 2000) but in studying the processing strategies using symbolic descriptions of objects and prototypes by graphs and a hierarchical organized *winner-takes-all* (WTA) net. This net will classify objects by competitive matching with a set of predefined prototypes in a *self-organizing manner*, i.e. without a *homunculus* acting as a supervisor. The investigation of the WTA-processing strategies might also shed light on principles of functioning of the Short-Term-Memory (e.g. Grossberg, 1987a; Grossberg, 1987b), on the role of attention (Lee et al., 1999), and on a trade-off between accuracy vs. speed of recognition depending on the strength of inhibition as shown in our first experimental results given below.

In Artificial Intelligence and Image Recognition graphs are a well suited representation of relational structures like molecular structures, data structures, or semantic networks. In any case, a relational structure consists of elementary objects and binary relations between these objects. In a graph of a relational structure the elementary objects are represented by vertices and their relations by directed or undirected edges. For example, in chemistry, graphs model molecular structures where the vertices represent atoms and the edges represent bonds. In Computer Vision vertices of a graph are objects within a scene and edges are structural relationships between those objects.

A fundamental problem in many application domains of processing relational structures is the identification and recognition of common structural parts between two relational structures. For example in classification, recognition or clustering tasks, information about structural overlaps between two structures is required in order to determine a similarity or distance of these structures. Here we call the computation of a similarity or distance between two relational structures *graph matching*.

In general graph matching problems are well-known NP-complete problems (Garey & Johnson, 1979). Due to the high computational complexity much effort has been directed toward devising efficient heuristics to find optimal or approximate solutions for graph matching problems. Among other heuristics artificial neural networks have been proposed as a promising model of computation for solving graph matching problems (Schädler & Wysotzki, 1999).

The high computational complexity is even more inconvenient if the solution of a problem requires several graph matching procedures. In distance-based classification using neural networks an input graph G is matched against a given set of N model graphs $M_1, \ldots, M_N$ representing prototypes of category $C_1, \ldots, C_N$, respectively. The matching is performed by recurrent neural networks $S_1, \ldots, S_N$. In

the following we will call these networks S_k *subnets*.

One fundamental approach in distance-based classification of structures is the classification by means of a *discriminant function* and a *comparative maximum selector* (CMS) classifier. A classification task is solved by a CMS approach in the following chronological order (Schädler & Wysotzki, 1999): (1) A feature extractor computes the discriminants ϱ_k of input G and each model M_k. The discriminant ϱ_k is computed by the k-th subnet S_k and serves as a measure for the similarity between G and M_k. (2) The discriminants are passed to a CMS classifier. (3) The CMS classifier sequentially compares the calculated discriminants and assigns the input graph to the category for which the corresponding discriminant is largest. Thus a CMS classifier processes its incoming data like a supervising *homunculus*.

In the traditional CMS classifier approach each match between G and M_k has equal priority in the sense, that each subnet S_k evolves until it has computed a discriminant ϱ_k for G and M_k. Thus a CMS classifier completes the evolution of all N subnets although the internal states of some subnets may indicate high dissimilarity of the graphs to be compared at an *early stage* during the matching process.

In order to improve the computational performance of a classifier for relational structures and to investigate the effects of self-organization in hierarchical networks we present a *self-organizing structure recognition* (SOSR) network. To accomplish a better computational performance than the CMS classifier the SOSR model identifies dissimilar pairs G and M_k at an early stage of the matching process and aborts the computation of the corresponding subnets S_k. Thus the SOSR network focuses on promising subnets and neglects subnets indicating high dissimilarity between the input and the corresponding model. This mechanism is realized by replacing the CMS by an inhibitory WTA network for the maximum selection and intertwining the tasks of step (1)-(3). Further improvements to reduce the computational effort can be made by parallel processing which is not possible with a sequential CMS classifier.

Note, that the SOSR model has much in common with the *competitive relational mind model* proposed by Taylor (1996). Furthermore, Grossberg (1987a) uses in a similar way a network for competitive learning with a reset mechanism which deactivates subnets announcing high dissimilarity to allow them to rebuild the model, i.e. the expectation.

The rest of this paper is organized as follows: We conclude this introductory section with some basic notations and definitions. In the next Section we formulate the graph matching problem in terms of the maximum clique problem. Subsequently we describe the SOSR network architecture. In an empirical study we investigate the behavior of the SOSR system. Finally, the last section summarizes this

contribution.

Notations and Definitions: A *graph* is a pair $G = (V, E)$ consisting of a finite set $V \neq \emptyset$ of *nodes* and a binary relation $E \subseteq V^2 := \{(i, j) \mid i, j \in V, \ i \neq j\}$. The elements $(i, j) \in E$ are called *edges*. A *subgraph* $H = (V_H, E_H)$ of G is a graph with $V_H \subseteq V$ and $E_H \subseteq V_H^2 \cap E$. An *induced subgraph* H of G is a subgraph with $E_H = V_H^2 \cap E$. A graph G is called *complete*, if $E = V^2$. A complete subgraph $C \subseteq G$ is called *clique* of G. A *maximum clique* $C \subseteq G$ is a clique with maximum number of vertices. A *maximal clique* is a clique which is not contained in any larger clique. The *clique number* $\omega(G)$ of a graph G is the number of vertices of a maximum clique in G. The *size* n_V of a graph G is the number $|V|$ of its vertices.

Graph Matching and Maximal Cliques

The graph matching problem is the problem to find the best partial mapping between two graphs where the quality of the mapping is estimated in terms of a problem dependent objective function. Our SOSR approach can also be applied to inexact graph matching problems of colored graphs, where vertex colors represent elementary objects and edges colors represent the type of relation between theses objects. But for convenience we only consider the common maximum induced subgraph problem which comprises the graph isomorphism and subgraph isomorphism problem as special cases.

A common approach to solve these classes of graph matching problems is based on maximum clique detection in an *association graph* (Ballard & Brown, 1982)). The association graph is formed by creating vertices $\boldsymbol{i}$ from each pair of vertices $(i_1, i_2) \in V_1 \times V_2$ and inserting edges between vertices $\boldsymbol{i} = (i_1, i_2)$ and $\boldsymbol{j} = (j_1, j_2)$ if (i_1, j_1) and (i_2, j_2) are edges in G_1 and G_2, respectively.

By definition of an association graph the cliques of $A = A(G_1, G_2)$ are in 1-1 correspondence to common isomorphic induced subgraphs of G_1 and G_2 and the maximum cliques uniquely correspond to the common maximum induced subgraphs of G_1 and G_2. This maps the graph matching problem to the optimization problem of finding a maximum clique C in A where the discriminant $\varrho(A)$ is a function on the number of vertices of C.

The SOSR Model

In the following let $A_k = (V_k, E_k)$ be the association graph of input G and model M_k ($1 \leq k \leq N$), where M_k represents category C_k. With ϱ_k we denote the discriminant of A_k.

The SOSR model consists of two interconnected layers, a feature extractor layer and a classifier layer (see Figure 1). The feature layer contains N subnets S_k each comparing input G with model M_k. The

classifier layer is a competitive WTA network for the maximum selection consisting of N inhibitory connected units where unit c_k represents category C_k. Each subnet S_k of the feature extractor is connected to unit c_k of the WTA classifier via an *inter-unit* i_k.

classifier layer

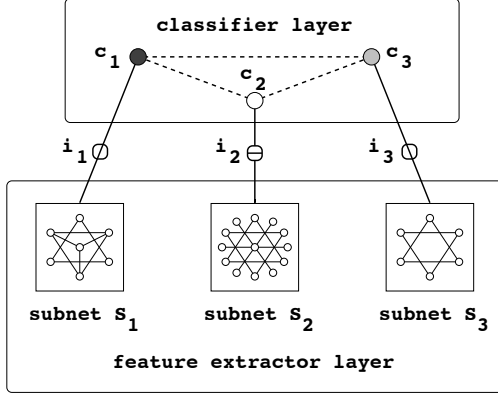subnet S_1 subnet S_2 subnet S_3

feature extractor layer

Figure 1: Architecture of a SOSR network.

During classification the subnets S_k evolve synchronously and continuously pass their current internal states to the inter-units i_k. The inter-units compute interim values $\varrho_k(t)$ of the discriminants ϱ_k and transfer them to the WTA classifier. The WTA classifier evaluates the evidence presented for a decision at an early stage. If the activation $z_k(t)$ of unit c_k in the classifier layer falls below zero, the WTA network disconnects unit c_k from inter-unit i_k, such that subnet S_k is excluded from the competition.

Figure 1 depicts a functional diagram of the SOSR model for the $N = 3$ category problem. The shading of units c_k in the classifier layer indicates their output level where darker shading means a higher output. Thus unit c_1 is dominating while unit c_2 has the lowest activation. Inter-units i_k are depicted as switches. A subnet S_k participates in the competition if the connection between S_k and c_k is switched on. Otherwise, S_k is considered to be irrelevant for the decision making process and inter-unit i_k switches off the connection to exclude S_k from the competition. In figure 1 inter-unit i_2 disconnects subnet S_2 from the classifier layer.

In mathematical terms the following equations describe the behavior of the SOSR model:

$$\dot{x}_i^k(t) = -dx_i^k(t) + \sum_j w_{ij} f\left(x_j^k(t)\right) \quad (1)$$

$$y_k(t) = \left[\left[\varrho\left(\boldsymbol{x_k}(t)\right)\right]_0^1 + \varepsilon - s_\varepsilon\left(-z_k(t)\right)\right]_0^1 \quad (2)$$

$$\dot{z}_k(t) = -d' z_k(t) - w \sum_{j \neq k} \left[z_j(t)\right]_0 + y_k(t) \quad (3)$$

where $[x]_\theta^\vartheta := \max\{\theta, \min\{x, \vartheta\}\}$ is the limiter function with lower and upper bound $\theta < \vartheta$, $[x]_\theta :=$

$\max\{\theta, x\}$ is the linear threshold function with lower bound θ, and s_ε is a trigger function of the form

$$s_\varepsilon(x) = \begin{cases} 1 + \varepsilon & : \quad x \geq 0 \\ 0 & : \quad x < 0 \end{cases}$$

Equation (1) describes the dynamics of an additive recurrent subnet S_k in the feature extractor, equation (2) describes the behavior of inter-unit i_k connecting subnet S_k with unit c_k of the WTA net, and equation (3) specifies the WTA dynamics. The system terminates if only a single unit c_{k_*} in the WTA classifier is left with an activation $z_{k_*}(t) > 0$ while all other units c_k are inhibited, i.e. $z_k(t) \leq 0$. Under the assumption of converging subnets termination follows from (Jain & Wysotzki, 2001a; Wersing & Beyn & Ritter, 2001).

Equation (1): A Maximal Clique Solver

Let index k refer to subnet S_k solving the maximum clique problem for graph A_k.

Many different neural network approaches and techniques have been proposed to solve the maximum clique problem (Bomze et al., 1999). As a representative model we consider a general additive recurrent network (1) where $x_i^k(t)$ denotes the activity of unit i of subnet k at time t and the constant $d \in [0, 1]$ describes the self-inhibition. The strength of the connection between unit i and unit j is determined by the synaptic weight $w_{ij} = w_{ji}$. The output of each unit is computed by a non-decreasing bounded function f.

In order to solve the maximum clique problem of the k-th association graph $A_k = (V_k, E_k)$ the network consists of $|V_k| = n_k$ units which are connected with weight $w_{ij}^k = w_E^k > 0$ if $(i, j) \in E_k$ is an edge in A_k and with weight $w_{ij}^k = -w_I^k < 0$ if $(i, j) \notin E_k$. Self weights w_{ii}^k are set to zero.

Given appropriate parameter settings the maximal clique solver operates as follows (Schädler & Wysotzki, 1998; Schädler & Wysotzki, 1999): An initial activation is imposed on the network. Finding a maximum clique then proceeds in accordance with equation (1) until the system reaches a stable state. The stable states correspond to the maximal cliques of A_k. In the ideal case a maximum clique is found. The clique size can be read out by counting the number of units with activation $x_i^k(t) \geq 1$.

For our experiments we used the time-discrete approximation of (1) given by

$$x_i(t+1) = (1-d)x_i(t) + \sum_{j \neq i} w_{ij} f_T\left(x_j(t)\right) + \eta(t) \quad (4)$$

where $\eta(t)$ is a small random noise to dissolve ambiguities and f_T is a controllable limiter function of the form

$$f_T(x) = \begin{cases} 1 & : \quad x \geq T \\ 0 & : \quad x \leq 0 \\ x/T & : \quad \text{otherwise} \end{cases}$$

with control parameter T which we call the *pseudo-temperature*. Starting with a high inititial value for $T = T_0$ the pseudo-temperature T is decreased during the evolution of the network according to an annealing schedule until it reaches a sufficient low final value $T = T_f$. Applying an annealing scheme avoids that the system gets stuck in spurious minima. The annealing schedule is of the following form:

1. Intialize $T \leftarrow T_0$.

2. Let the network iterate τ steps according to the dynamical rule given in (4).

3. Decrease the pseudo-temperature by $T \leftarrow a \cdot T$ where $0 < a < 1$ is an annealing constant.

4. If $T > T_f$ continue with step 2. Otherwise terminate the algorithm.

The general parameter settings of the subnets S_k follows a theoretical analysis given in Jain & Wysotzki (2002). For the weights we set

$$w_E^k = \frac{2}{\deg_I^k \cdot (n_k - \deg_E^k)}$$
$$w_I^k = \deg_E^k \cdot w_E^k$$

where n_k is the number of units of subnet S_k and $\deg_E^k$ ($\deg_I^k$) is the maximum number of excitatory (inhibitory) connections of an unit in S_k.

Equation (2): Inter-Units:

An inter-unit i_k connects subnet $S_k = (V_k, E_k)$ with unit c_k in the WTA classifier and controls the external input $y_k(t)$ of c_k.

Let $\boldsymbol{x_k}(t)$ be the current state vector of subnet S_k. Inter-unit i_k receives $\boldsymbol{x_k}(t)$ as its input and computes an interim value $\varrho_k(t) = \varrho(\boldsymbol{x_k}(t))$ of the discriminant ϱ_k. The computation of interim values $\varrho_k(t)$ is constrained to

$$\varrho_k(t) \leq \varrho_k \qquad (5)$$

where equality holds if and only if S_k is in a stable state corresponding to a maximum clique. The discrimination value ϱ_k measures the resemblance between input G and model M_k. Thus the interim values $\varrho_k(t)$ reflect a preliminary estimate of the discriminant ϱ_k where the degree of resemblance of G and M_k gradually emerges with the time spent on the matching problem. This is indicated by (5). At beginning an interim value $\varrho_k(t)$ is at a low level and with increasing time $\varrho_k(t)$ approaches ϱ_k. During evolution of S_k it is not required that $\varrho_k(t)$ is monotonously increasing with the time. We call local maxima and minima of $\varrho_k(t)$ *deceptions*. Deceptions may leed to misclassifications. A local minimum of $\varrho_k(t)$ may result in an exclusion of S_k from the competition. In this case the resemblance of G and M_k is underestimated during the comparision

of the structures G and M_k. Similarly, a local maximum of $\varrho_k(t)$ may result in a premature decision which assigns G to category C_k. Here the match of input G and model M_k is overestimated during computation. Another source of misclassifications arise from insufficient synchronization among the evolving neural maximal clique solvers S_k. Here, too fast (too slow) convergence of $\varrho_k(t)$ to a low (high) discriminant ϱ_k can lead to an erroneous decision. Thus it is an important objective to design the computation of $\varrho_k(t)$, such that deceptions are avoided and the matching procedures of the subnets S_k are synchronized.

Depending on its current state $y_k(t)$ an inter-unit i_k transfers the interim value $\varrho_k(t)$ to unit c_k of the WTA classifier. An inter-unit i_k is in state ON if $y_k(t) > 0$ and in state OFF if $y_k(t) \leq 0$. Only inter-units in state ON pass interim values to the WTA classifier. Initially, the state of each inter-unit is ON where $\varepsilon > 0$ is a small constant to avoid a deactivation of i_k at the beginning. According to (3) let $z_k(t)$ be the activation of unit c_k in the WTA classifier. Unit c_k switches OFF inter-unit i_k if $z_k(t) \leq 0$ and does not effect i_k otherwise. This mechanism is realized by the trigger function s_ε. If $z_k(t) \leq 0$ then $-z_k(t) \geq 0$ and thus the value $s_\varepsilon(-z_k(t)) = 1 + \varepsilon$ is subtracted from $\phi_k(t) + \varepsilon \leq 1 + \varepsilon$. This sets the activation level $y_k(t)$ of inter-unit i_k equal to 0. Similarly, if $z_k(t) > 0$ the trigger function s_ε has no influence on the activation $y_k(t)$ of inter-unit i_k. Once in state OFF an inter-unit can not be reactivated. In practical applications the corresponding subnets S_k can be switched off.

Next we give an example how to compute the interim values $\varrho_k(t)$ on the basis of a family of discriminant functions

$$\varrho_k = \frac{\alpha|V_{C_k}| + \beta|E_{C_k}|}{\mu_k}$$

where $\alpha, \beta \geq 0$ are problem specific constants which weight the vertex and edge matches, $C_k = (V_{C_k}, E_{C_k})$ is a maximum clique of A_k, and $\mu_k > 0$ is a normalization constant to ensure an upper bound of 1. Note, that $1 - \varrho_k$ defines a family of distance metrics (Schädler & Wysotzki, 1999). We define $\varrho_k(t)$ to be a function on the number of θ-*active units*:

1. Let $V_k^\theta \subseteq V_k$ be the set of θ-active units with activation $x_i^k(t) \geq \theta$ where $0 \leq \theta \leq 1$ is a threshold. Compute the current θ-*intensity*

$$\sigma_V^\theta(t) = \sum_{i \in V_k^\theta} \left[x_i^k(t)\right]_\theta^1 \qquad (6)$$

2. Let E_k^θ be the set of all excitatory connections (i, j) between θ-active units $i, j \in V_k^\theta$. Compute

the current θ-*connective-intensity*

$$\sigma_E^\theta(t) = \sum_{(i,j) \in E_k^\theta} \left[x_i^k(t) \right]_\theta^1 + \left[x_j^k(t) \right]_\theta^1 \qquad (7)$$

3. Let I_k^θ be the set of all inhibitory connections (i,j) between θ-active units $i, j \in V_k^\theta$. Compute the current θ-*incompatibility*

$$\sigma_I^\theta(t) = \sum_{(i,j) \in I_k^\theta} \pi_{ij}^k(t) \qquad (8)$$

where $\pi_{ij}^k(t) \geq 0$ is a penalty term for the presence of inhibitory connected active units.

4. Compute the current interim value

$$\varrho_k(t) = \frac{\alpha \sigma_V^\theta(t) + \beta \sigma_E^\theta(t)}{\mu_k} - \sigma_I^\theta(t) \qquad (9)$$

as a function of the current θ-intensity, the current θ-connective-intensity, and the current θ-incompatibility.

The appropriate choice of $\pi_{ij}^k(t)$ is crucial to synchronize the subnets and to avoid deceptions. Too high or too low values of $\pi_{ij}^k(t)$ result in a higher percentage of misclassifications.

Note, that in a stable state of S_k the interim values $\varrho_k(t)$ are identical to the discriminant ϱ_k defined by a corresponding maximal clique C_k.

Equation (3): A WTA Classifier:

The dynamical system given by equation (3) is an inhibitory WTA network for the maximum selection from a set of external input signals where unit k represents category C_k, $z_k(t)$ is the activation of unit c_k, $-w < 0$ represents the inhibitory strength of the synapses connecting any pair of units, $d' > 0$ is the self-inhibition, and $y_k(t)$ is the external input from inter-unit i_k.

Given an initial input vector the operation of these networks is a mode of contrast adjustment and pattern normalization where only the unit with the highest activation fires and all other units in the network are inhibited after some setting time (Jain & Wysotzki, 2001a; Jain & Wysotzki, 2001b).

In the context of the relational mind model of Taylor (1996) the WTA classifier may be seen as the analogue of the thalamic NRT-C complex acting as a central executive for global competition.

Experiments

In first experiments we focused on the performance of the proposed SOSR model and on the role of the inhibition $-w$.

To keep the experiments simple we considered the more general task of identifying a graph A_{k_0} with

n_v / t_{MS}	25 / 180.1		50 / 341.3	
w	κ	t_{SO}	κ	t_{SO}
0.2	1.00	18.8	0.95	33.0
0.1	1.00	19.2	0.95	44.0
0.01	1.00	75.2	0.99	97.5
0.001	0.99	167.3	0.94	251.1
0.0001	0.99	178.7	0.85	323.7

Table 1: Classification accuracies κ and number of iterations t_{SO} for size $n_V = 25$ and $n_V = 50$.

maximal clique number ω_{k_0} among a set of $N = 10$ graphs $A_1, \ldots, A_N$ of identical size $n_V := |V_1| = \cdots = |V_N|$. For a single run of the algorithm with fixed size n_V the clique numbers $\omega_1, \ldots, \omega_N$ of the N graphs were drawn from a Gaussian distribution restricted to the interval $[3, n_V - 1]$ with identically distributed random mean $3 \leq E[\omega] < n_V$ and variance $0 < Var[\omega] < n_V/2$. The chosen sizes are $n_V = \{25, 50, 100, 250\}$. To assess the effects of the inhibition $-w$ we varied the weight $w = \{20^{-1}, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$. We performed 100 runs of the SOSR algorithm for each chosen n_V and w.

Parameter Settings: We have chosen $T_0 = w_E \cdot \deg_E^2 + w_I \cdot \deg_I^2$, $T_f = 0.1$, $a = 0.75$, and $\tau = 0.6 \cdot n_V$ as parameters for the annealing schedule of the subnets S_k. The self-inhibitions d and d' of the subnets S_k and the WTA classifier are set to 0.

To compute the discriminants we set $\alpha = 0$, $\beta = 1$, $\theta = 0$, $\varepsilon = 0.1$, and $\mu_k = \varepsilon^{-1} \cdot \omega_{k_0}^2$ for all k where ω_{k_0} is the maximal clique number of a sample $\omega_1, \ldots, \omega_N$. The θ-incompatibility $\sigma_I^\theta(t)$ is defined by the heuristically chosen penalties $\pi_{ij}^k(t) = \frac{4}{n_V} \left([x_i^k(t)]_\theta^1 + [x_j^k(t)]_\theta^1 \right)$.

Results: Table 1 and 2 summarizes the results. The first row shows the size n_V and the average iterations t_{MS} of a subnet *without switching* OFF, averaged over all N subnets and all 500 runs for size n_V. The other entries show for each size n_V the rate of correct classifications κ and the average iterations t_{SO} of a subnet *with switching* OFF, averaged over all N subnets and all 100 runs for a given weight $-w$ and size n_V. For example a subnet S_k consisting of 150 units in a 10-category SOSR system with $w = 0.01$ averages 230.8 iterations until it terminates. In a CMS classifier system S_k averages 993.7 iterations. The classification accuracy of the SOSR system for this configuration is 0.92.

Discussion: If the inhibition $-w$ in the WTA classifier is on a low level an increase will tend to more accurate and faster generated responses. But when

n_v / t_{MS}	100 / 571.0		150 / 993.7	
w	κ	t_{SO}	κ	t_{SO}
0.2	0.89	45.5	0.83	140.0
0.1	0.94	49.9	0.84	130.3
0.01	0.94	126.4	0.92	230.8
0.001	0.87	361.9	0.92	604.3
0.0001	0.74	543.7	0.67	882.0

Table 2: Classification accuracies κ and number of iterations t_{SO} for size $n_V = 100$ and $n_V = 150$.

inhibition is at a high level, the increased competition may interfere with correct decisions by trapping into deceptions or by immobility in finding any response. Thus there will be an optimal level of inhibition for effective behavior. The optimal setting of w compromises the conflicting tasks of significantly improving the computational efforts and gaining high classification accuracy.

Conclusion

We introduced a new self-organizing structure recognition system. The architecture couples the subnets in the feature extractor to a WTA classifier such that the subnets are in an indirect competition during the matching process. The system consecutively switches off subnets if their interim values indicate a *worse* match than the remaining active subnets and shifts its attention to more promising subnets. In first experiments we showed that the switching-off mechanism of the SOSR system significantly reduces the computational effort without suffering in substantial losses of classification accuracy. The inhibition $-w$ of the WTA classifier controls the conflicting interests of high classification accuracy and fast decision making.

Ballard, D.H. & Brown, C.M. (1982). *Computer Vision*. Englewood Cliffs, NJ: Prentice Hall.

Bomze, I.M. & Budinich, M. & Pardalos, P.M. & Pelillo, M. (1999). The maximum clique problem. In D.-Z. Du & P.M. Pardalos (Eds.), *Handbook of Combinatorial Optimization*. Boston, MA: Kluwer Academic Publishers.

Garey, M. & Johnson, D. (1979).*Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York: W.H. Freeman and Company.

Grossberg, S. (1987a). Competitive Learning: From interactive activation to adaptive resonance. *Cognitive Science, 11*, 23-63.

Grossberg, S. (1987b). *The Adaptive Brain*. Elsevier: Amsterdam.

Hinton, G. (1994). Wie Neuronale Netze aus Erfahrung lernen. In W. Singer (Ed.), *Gehirn und Bewusstsein*. Spektrum.

Jain, B.J. & Wysotzki, F. (2001a). On the short-term-memory of WTA nets. In M. Verleysen(Ed.), *Proceedings of the Ninth European Symposium on Artificial Neural Networks* (pp. 289-294). Brussels, Belgium: D-Facto.

Jain, B.J. & Wysotzki, F. (2001b). Efficient Pattern Discrimination with Inhibitory WTA Nets. In G. Dorffner, H. Bischof, K. Hornik (Eds.), *Artificial Neural Networks - ICANN 2001* (pp. 827-834). Springer.

Jain, B.J. & Wysotzki, F. (2002). Parameter Settings of a Neural Network for the Maximum Clique Problem. *To appear*.

Lee, D. & Itti, L. & Koch, C. & Braun, J. (1999). Attention activates winner-take-all competition among visual filter. *Neural Neuroscience, 2*, 375-381.

Rosch, E. (1975). Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General, 104*, 192-233.

Rosch, E. & Lloyd, B. (1978). *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schädler, K. & Wysotzki, F. (1998). Application of a neural net in classification & knowledge discovery. In M. Verleysen(Ed.), *Proceedings of the sixth European Symposium on Artificial Neural Networks* (pp. 117-122). Brussels, Belgium: D-Facto.

Schädler, K. & Wysotzki, F. (1999). Comparing Structures Using a Hopfield-Style Neural Network. *Applied Intelligence, 11*, 15-30.

Singer, W. (2000). Response Synchronization: A Universal Coding Strategy for the Definition of Relations. In M.S. Gazzaniga (Ed), *The New Cognitive Neurosciences*. MIT Press.

Taylor, J.G. (1993). A Global Gating Model of Attention and Consciousness. In M. Oaksford & G. Brown (Eds.), *Neuro-Dynamics and Psychology*. New York: Academic Press.

Taylor, J.G. (1996). Modelling what is like to be. In S.R. Hameroff et al. (Eds.), *Toward a Science of Consciousness*. MIT Press.

Wersing, H. & Beyn, W.-J. & Ritter, H. (2001). Dynamical Stability Conditions for Recurrent Neural Networks with Unsaturating Piecewise Linear Transfer Functions. *Neural Computation, 13*, 1811-1825.