On the Role of Hierarchy for Neural Network Interpretation

Jürgen Rahmel†, Christian Blum†, Peter Hahn‡
†University of Kaiserslautern, Centre for Learning Systems and Applications
PO Box 3049, 67653 Kaiserslautern, Germany
‡Neustadt Hand Centre
Salzburger Leite 1, 97616 Bad Neustadt, Germany

Abstract

In this paper, we concentrate on the expressive power of hierarchical structures in neural networks. Recently, the so-called SplitNet model was introduced. It develops a dynamic network structure based on growing and splitting Kohonen chains and it belongs to the class of topology preserving networks. We briefly introduce the basics of this model and explain the different sources of information built up during the training phase, namely the neuron distribution, the final topology of the network, and the emerging hierarchical structure. In contrast to most other neural models in which the structure is only a means to get desired results, in SplitNet the structure itself is part of the aim. Our focus then lies on the interpretation of the hierarchy produced by the training algorithm and we relate our findings to a common data analysis method, the hierarchical cluster analysis. We illustrate the results of network application to a real medical diagnosis and monitoring task in the domain of nerve lesions of the human hand.

1 Introduction

Existing approaches to hierarchical clustering and classification neglect the spatial relations of clusters or partial solution spaces. A particular class of neural network models has the potential to overcome this problem. Regarding the mapping from the input space onto the space spanned by the neighborhood relations of the neurons in the network, the property of certain neural network models to keep track of neighborhood relationships of clusters of data even in cases of reduction of dimensionality is called *topology preservation*. The degree of topology preservation can be determined by the observation, how well neighborhood relationships in one space are preserved by the mapping onto the other space. Thus, one question is: for two input vectors that are close in input space, are their best matching units close[1] in the network topology? The other question is: for two neurons that are neighbors in the network topology, are their associated weight vectors close in the input space? These questions led to the development of the topographic function [Villmann *et al.*, 1996], that effectively quantifies the topology preservation in topographic maps. We call locations, where those maps are not continuous *topological defects*. Topology preserving models are, among others, the Self-Organizing Map (SOM) [Kohonen, 1990], the Growing Cell Structures (GCS) [Fritzke, 1993] and the Topology Representing Network (TRN) [Martinetz and Schulten, 1994] as well as several descandants of these examples. But all those models lack the ability of hierarchically structuring the training set. Depending on the reduction of dimensionality performed by the models, there is a principal difference in degree of topology preservation each model can achieve. If the dimension of the network structure is limited, as it is the case for the SOM and the GCS, the real dimensionality of the data space may necessarily invoke topological defects.

The SplitNet model for the first time succeeded in developing a hierarchical structure over the training set. It is an unsupervised learning method and in some sense comparable to the hierarchical cluster analysis (see e.g. [Duda and Hart, 1973]). For static models, like e.g. the Self-Organizing Map, the task of the network application determines the desired interpretability of the network and thus controls such parameters of network design as number and connectivity of neurons. In dynamically growing networks, the approach is necessarily different. The incremental construction of the network up to its final size and topology is in general controlled by specific performance criteria. The training result is a network, where not only the weights contain relevant information. Additionally, the size of the network, the distribution of neurons and the emerged connectivity inside the network implicitly encode information on the trained sample set. Compared to the above-mentioned topology preserving models, the tree-structured organization of topologically connected parts of the SplitNet network adds a completely new dimension to the interpretability of the network model. The hierarchy offers

[1] As different input vectors may be mapped onto the same neuron, in this informal explanation, the *closeness* of neurons includes also the identity of those neurons.

structured knowledge on various levels of abstraction and generalization as well as optimized access to samples of the training population.

The rest of the paper is organized as follows. In the next section, we will outline the basic methods related to the neural model used in our approach. Section 3 will present the principle of the SplitNet model and in Sec. 4 the role of the emerging hierarchical structure is explained. We then present results of the application of the SplitNet model in the medical domain of finger movement pattern analysis. A summary and final remarks will conclude the paper.

2 Related Work

The hierarchical cluster analysis, either the divisive or the agglomerative approaches, are methods that progressively split or link clusters of data. The result of the analysis can be visualized as a dendrogram (see Fig. 1), which is a two-dimensional tree structure that shows the order of linkage (for the agglomerative case) and the distance or similarity at which this linkage of clusters was performed. Thus, this method is able to display the clustering of data, whereby the result depends on the distance measure that determines the closeness of clusters and/or samples. Specific variants of agglomerative versions of the hierarchical cluster analysis are e.g. the *single linkage* and *complete linkage* methods, which minimize the minimum or maximum distance of cluster elements, respectively, during each merger of two clusters. For the comparison intended in this paper (cf. Sec. 5), we will use the *centroid method*, which selects those clusters with the minimum distance between their means.

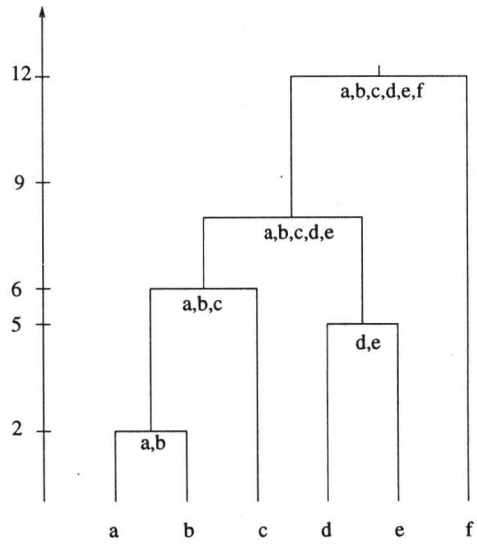

Figure 1: Example of a dendrogram and the distance information available for each linkage level

Because of the explicit distance information contained in the dendrogram for each linkage of clusters, hierarchical clustering is a flexible way of detecting the resulting number of clusters given a certain threshold value. However, it is not possible to reason from the real spatial relationship of the observed pattern. There is no similarity information other than the one for linked clusters. Similar statements are true of course for divisive methods. So the hierarchical clustering methods are useful tools for a preliminary analysis of the data, but they do not provide additional ways for explanation of the clustering results and do not enable reasoning on alternative solutions based on neighborhood observations.

Such inspection of neighboring clusters and samples can be performed by topology preserving networks. As indicated above, several models like the GCS or the TRN already exist for topology preserving representation of a training set. The Growing Cell Structures (GCS) are a dynamic vector quantization model. Different criteria, e.g. the quantization error, determine the insertion position of a new neuron. Removal strategies yield an adaptive quantizer that is superior to the original SOM, but the GCS model also uses an a priori specified dimensionality (through the choice of simplices) and thus is prone to the appearance of topological defects for high dimensional data spaces. The TRN algorithm also approximates the distribution of input data and constructs topology preserving connections between its neurons. In the limit, it is able to find the Delaunay triangulation of a data set, thus it generates, by virtue of not being fixed to a given dimensionality, a nearly perfectly topology preserving map. But both neural models only provide data analysis on a flat level. They cannot provide views on the data at different granularities, and thus lack the advantages of methods like the hierarchical cluster analysis.

3 The SplitNet Model

SplitNet is a topology preserving, dynamically growing model for unsupervised learning and hierarchical structuring of data [Rahmel, 1996b]. Starting with a single, small Kohonen chain [Kohonen, 1990], localized insertion and deletion criteria enable an efficient quantization of the data space. The hierarchy in the architecture grows, if one of the following splitting criteria is satisfied:

- detection of topological defects
- deletion of neurons by an aging mechanism
- significant local variances in quantization errors or
- significant local variances in edge lengths.

Those criteria are checked several times during progress of training. If a criterion is satisfied, the affected chain is split into two or more subchains which are added to the network at one level lower in the hierarchy. The node in the hierarchy that formerly represented the unsplit chain now serves as a generalized description and access structure for the new son nodes. Figure 2 illustrates this basic mechanism. If a topological defect is found (e.g. between neurons 1 and 7, which are close in input space but distant in the chain of neurons), the chain is split and

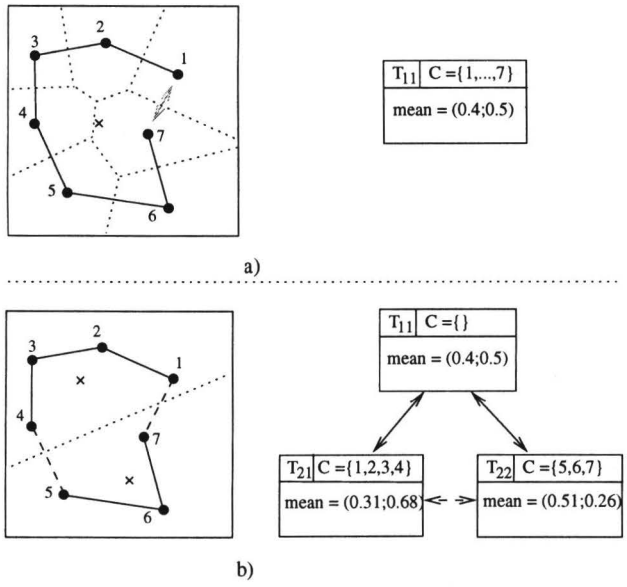

Figure 2: Example of splitting a chain because of a topological defect

nodes representing the fragments are added as descendants to the tree. Path decisions in the so constructed hierarchy will be made according to the mean of the weight vectors of the neurons in the chain, therefore the mean is also indicated in the figure. The topology preserving construction of the network structure provides local neighborhood information that is necessary for incremental retrieval of nearest neighbor to a given input vector. The dashed lines in Fig. 2 indicate this type of knowledge. The neighborhood relations are kept as lateral connections defining the topology of the network space. They are responsible for the high degree of topology preservation in SplitNet and enable a fast and incremental search for a set of nearest neighbors. A more rigorous and exhaustive treatment of these aspects and retrieval results can be found in [Rahmel and Villmann, 1996]. The purpose of such a set of nearest neighbors is e.g. to serve as the basis for the k-nearest-neighbor rule in decision making processes (cf. Sec. 5).

Since unsupervised learning methods provide no direct classification, the training result has to be interpreted in the context given by the training data. For the SplitNet model, we observe three containers of knowledge that can be used for the tasks in applications like the one described in this paper:

Neuron distribution: The insertion criterion determines the error function to be minimized. Quantization of the data set allows local estimation of sample density.

Topology: The connections between neighboring neurons provide information on where to find similar cases. Measuring topological defects yields the search depth for incremental retrieval of nearest neighbors to a given query.

Hierarchy: The hierarchical structure of the network contains different levels of generalization and abstraction. It allows a fast tree search for best matches and insightful visualization of the data structure for the domain expert.

The utility of the neuron distribution is comparable to reference vector placement in quantization algorithms [Gray, 1984]. The neuron positions thus minimize the average reconstruction error for all elements of the data set. Interpretation of the network topology is described e.g. in [Rahmel, 1996a]. It can be shown that SplitNet produces networks with only small topological defects, indicated by low values of the topographic function [Villmann *et al.*, 1996]. This specific property limits the search effort for procedures like the probing algorithm [Lampinnen and Oja, 1989], which conquers local neighborhoods of neurons in order to find a better match to a given input. In the following, we illustrate the semantics of the hierarchical structure developed by methods like SplitNet.

4 Interpretation of Hierarchy

Hierarchical organizations offer additional properties to make use of in data analysis and structure utilization. One rather general aspect is the fact that a hierarchical structure – like any search tree – provides fast access to the terminal nodes. For neural networks like SplitNet, this results in accelerated training runs since the search for the best matching unit is supported by the hierarchical network structure. This determination of the best match can be summarized as follows:

- tree search for a candidate unit using the means of chain vectors for descending the hierarchy, and then
- local search through topological connections for a possibly better match.

The local search costs additional time, but since it is strictly local and depending on the degree of topology preservation in the network, it could demonstrated that the improved topology preservation of the SplitNet model considerably increased the speed of best matching unit access [Rahmel, 1997]. In addition, the larger the network, the less the influence of the additional local search procedure in comparison to the savings due to the hierarchical arrangement.

The hierarchies produced by classification or decision trees [Breiman *et al.*, 1984] [Quinlan, 1993] yield simple, crisp, and explicit tests as path decisions in nodes at the expense of flexibility of the decision regions. The orientation of hyperplanes generated by those tests is limited to the dimensionality of the respective test. Unfortunately, higher dimensionality of the test yields both higher flexibility and massively growing computational effort for determination of optimal tests. Therefore, in practical applications, tests are often one- or two- dimensional and the corresponding hyperplanes separating subspaces of the sample space are orthogonal to the

coordinate axes or depending on only two of the possible vector components. In contrast to this, the SplitNet structure offers implicit tests that cover the whole information contained in the description of a sample. The decision regions of SplitNet approximate the Voronoi regions given by the sample population and thus minimize quantization errors imposed by generalization inside the regions.

Classification or decision trees select the tests for path decisions according to the gain criterion for classification of samples. In an unsupervised setting where class information is not available, efficient sample localization plays the most important role. In order to minimize the search effort, we need a test that maximizes the information about the location of the nearest sample. The Kohonen model provides a solution for this task. As demonstrated in [Ritter *et al.*, 1992], the weight adaptation of the algorithm leads to a discrete approximation of principal curves by Kohonen chains. If we divide the sample population according to the placement of neurons and recursively repeat this subset construction, we get a hierarchic structure that on every level optimizes the information on sample locations. Thus, for an average test sample, we have an efficient access to the best match of the training population. In this respect, the tree is interpretable as a decision tree, regarding the optimization of spatial information, but still trainable and adpatable to slight changes in the training data set. It thus combines interpretability and flexibility of symbolic and connectionist machine learning methods, respectively.

The interpretation of a hierarchical structure like the one generated by SplitNet combines the knowledge on the above-mentioned property of the Kohonen algorithm with the splitting reasons of the SplitNet model, which deviate from this principle. Path decisions in the SplitNet tree have definite semantics to be used when descending the tree and relating accessed clusters with others that are reachable through the topology of the network.

5 Diagnosis and Monitoring of Ulnaris lesions

We now briefly present an application of the SplitNet model in the domain of nerve lesions of the human hand. We will outline the general problem and describe the results obtained by using the hierarchical neural model.

The human hand is provided with the radial, median and ulnar nerve. The ulnar nerve provides sensory function for the small and ring finger and innervates the intrinsic muscles of the hand. These muscles are crucial in balancing and coordinating the flexor and extensor muscles, rendering possible fine movement such as grip and pinch. While assessing sensory function is feasible, objective analysis of motor function is quite difficult. Clinical investigation includes grip force measurement and recording of active and passive range of motion. Besides these factors, ulnar nerve dysfunction causes changes in coordination of the movement which cannot be measured by instruments.

In contrast to a normal, physiological movement pattern (Fig. 3(a)), the dynamic disorder 'rolling' describes the pathological flexion of the finger. This movement resembles the rolling of a carpet (Fig. 3(b)). As an effect, patients are not able to grasp an object because their fingers push it out of the palm. The dynamic disorder 'clawing' describes the hyperextension of the MP joint with flexion of the PIP and DIP joint[2] while the finger is in resting position (Fig. 3(c)). These descriptions are based on the experience of the examiner. Changes in quality and especially improvement of fine motor activities after nerve repair are difficult to detect and to quantify. If nerve repair fails, there are different operations to rebuild the movement pattern. In these cases, the outcome of surgery also cannot be quantified. Until now, there was no convenient measurement system to distinguish finger movement patterns.

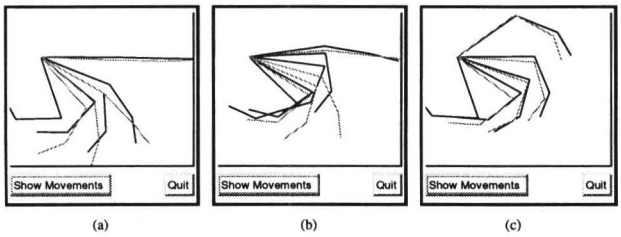

Figure 3: Different forms of finger movement pattern: (a) normal, physiological movement, (b) rolling, and (c) clawing (see text). Each picture shows nine steps of finger movement during one cycle of closing (black lines) and opening (gray lines) the fist.

Based on kinematic research we established a measurement system to get real-time data of human finger movement. Attempts to analyze these data with classical mathematical methods like discriminant analysis failed to distinguish between normal and pathological movement. Statistical clustering provides a good first insight into the structuring of the data but is not able to support the specific needs in this application like, for example, retrieval of samples and their comparison to a group of similar data, as it is required for diagnostic applications. For demonstration purposes, figure 4 shows an example of a tree structure generated with a small fraction of the available data. Despite the fact that our preprocessing generates high-dimensional training vectors, we used no further dimension reduction method. The reason is the necessity to display the hierarchy with the neuron weights retranslated into finger positions. By this, physicians can evaluate the position of a newly encountered data vector in the tree. In Fig. 4, the upper part of the tree contains roughly the patterns for 'clawing', while the bottom part corresponds to physiological

[2]The MP, PIP and DIP joints are the three finger joints ordered from the base joint between hand and finger to the tip.

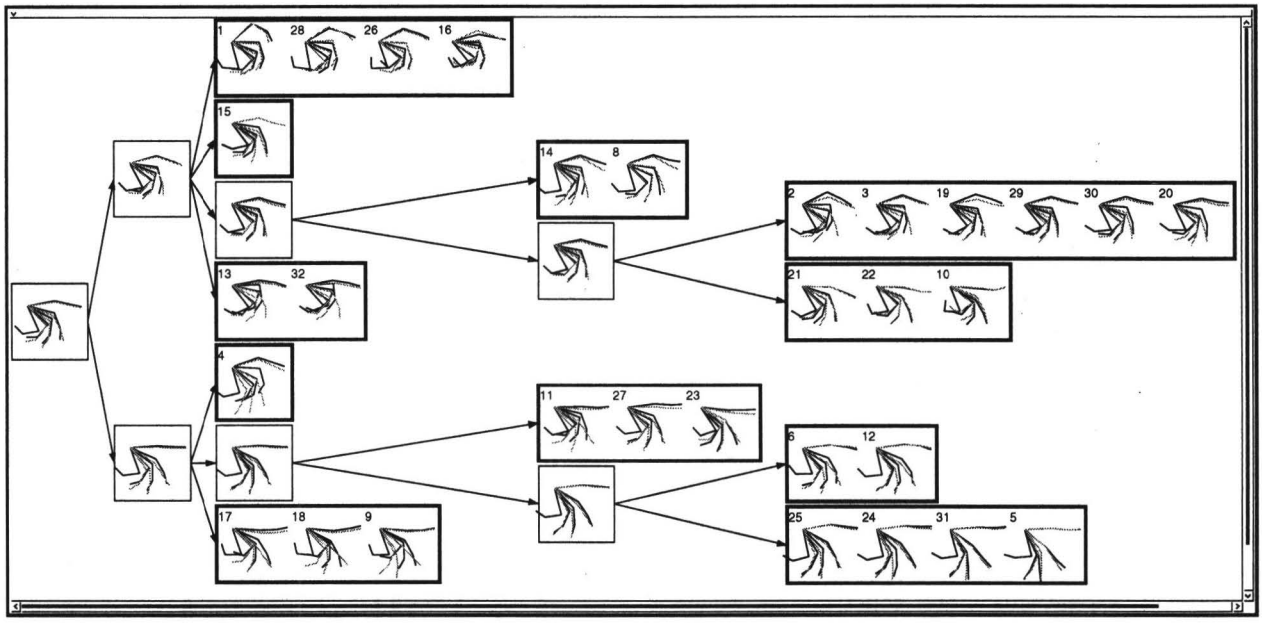

Figure 4: Hierarchical representations generated by SplitNet. The retranslation of neuron weights allows the display of interpretable finger movements in the learned hierarchical arrangement.

movement and 'rolling'.

Since we use unsupervised learning and therefore provide no class information for the training pattern, the resulting tree is obviously not a classification tree. The full information is here contained in the hierarchical structure **and** the topological connections of the nodes (which are not shown in the figure). Pattern # 4, the only 'clawing'-data in the lower subtree, is an example of the case that the division of the data space by the path decisions of the tree is suboptimal with respect to path decisions alone. Presentation of an input vector that closely matches the vector corresponding to pattern # 4 would cause the tree search (based on the vector means) to descend into the upper half of the tree in Fig. 4. Further path decisions would then yield pattern #14 as best match candidate. But the topological information that is accessible for the interpretation component of the network will show that this pattern is connected to pattern #4 in the lower subtree and local, topological search will identify the correct best match.

Similarly, further analysis of branches and subtrees reveals the full organization of the hierarchy. The sequence of generalizing non-terminal nodes supports the physician in understanding relative positions of different patient vectors. At each non-terminal node, the reason for branching – outlier splitting, topological defect, etc. – is accessible, so a reasonable interpretation of the emerging hierarchy, supported by the local topological connections (not shown in the figure), is rendered feasible.

We currently have more than 600 pattern of 55 patients with different forms of lesions and at various stages of motion recovery. Healing progress as well as success of surgery can be monitored in our network that is trained with all available data. A sequence of pattern representing the recovery process of a patient can be mapped onto the fully trained network. The interpretation component provides the physician with comparable cases for each pattern and so, by relating current data to previous cases, an evaluation of the actual healing process is possible.

A comparison of the results obtained by SplitNet with those of a hierarchical cluster analysis clarifies the strength of the neural model. We performed a run of the clustering process and examined a subgraph of the dendrogram with about as many terminal nodes as the SplitNet tree described above. The result was not surprising. The clustering produced nearly the same groups of data represented by leaf nodes, thus supporting the clustering abilities of the SplitNet model. However, despite the fact that distance information is available for the merging level of two clusters, interpretation of the dendrogram from a medical point of view was possible only in a very limited way. Whereas the neuron chains representing the terminal nodes in SplitNet arrange themselves in a direction that best reflects the largest variation in the associated movement pattern (the intrinsic property of the underlying Kohonen model), such information is not available in the cluster analysis. Moreover, the dendrogram provides information on the order of the cluster linkages, yet it does not contain explicit or implicit information on the spatial relationships of clusters. This is a crucial property for a reliable diagnosis of new cases which are not contained in the center of existing clusters. In order to compare those with nearest neighbors, reliable information on cluster connectivity is necessary.

The lateral connections between neurons in the SplitNet model facilitate reasoning for class assignment based on neighborhood considerations. We can use the retrieval properties of the topology preserving network structure for the enumeration of the nearest neighbors and application of the k-nearest-neighbor rule [Duda and Hart, 1973] yields a majority vote, if training samples can be associated with classification information.

6 Summary and Outlook

We briefly presented a recent neural network model which differs from existing models in the hierarchical structure that it creates by the training algorithm. The resulting network provides different sources of knowledge for network interpretation and we focused our discussion on the use of the hierarchy. We illustrated properties and advantages of the flexible tree structure. The applicability of the network model to real world tasks is shown with the example of a diagnosis system for ulnar nerve lesions. Our approach for the first time applies pattern recognition by a neural net approach to human finger movement. Besides simple clustering abilities, the applied SplitNet model provides support for the interpretation of both the learning processes which have occurred and the emerged hierarchical structure. Thus, in our case, interpretation of the images, which are retranslations of neuron weights into the semantics of training vectors, enhances our knowledge of the finger movement pattern.

So far, from the medical point of view, we do not know if we portray the whole spectrum of ulnar nerve dysfunction. More data have to be recorded and our aim is to build up a neural net containing all types of normal and pathological movement. Then we are able to represent all ulnar nerve lesions by recording finger movement and classify the new movement pattern by observing the mapping performed by the neural net onto a certain location in the tree, for which clinical diagnosis is already accessible.

References

[Breiman *et al.*, 1984] L. Breiman, J.H. Friedman, R.A. Olsen, and C.J. Stone. *Classification and Regression Trees.* Belmont, CA, Wadsworth, 1984.

[Duda and Hart, 1973] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis.* Wiley, 1973.

[Fritzke, 1993] B. Fritzke. Growing cell structures - a self-organizing network for unsupervised and supervised learning. Technical Report TR-93-026, ICSI, 1993.

[Gray, 1984] R. M. Gray. Vector quantization. *IEEE ASSP Magazine*, pages 4–29, April 1984.

[Kohonen, 1990] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

[Lampinnen and Oja, 1989] J. Lampinnen and E. Oja. Fast self-organization by the probing algorithm. In *Proc. of the IJCNN*, Washington, 1989.

[Martinetz and Schulten, 1994] Th. Martinetz and K. Schulten. Topology representing networks. *Neural Networks*, 7(2), 1994.

[Quinlan, 1993] J.R. Quinlan. *C4.5: Programs for Machine Learning.* Morgan Kaufman, 1993.

[Rahmel and Villmann, 1996] J. Rahmel and T. Villmann. Interpreting Topology Preserving Networks. Technical Report LSA-96-01E, University of Kaiserslautern, 1996.

[Rahmel, 1996a] J. Rahmel. On the Role of Topology for Neural Network Interpretation. In W. Wahlster, editor, *Proc. of the ECAI*, 1996.

[Rahmel, 1996b] J. Rahmel. SplitNet: Learning of Hierarchical Kohonen Chains. In *Proc. of the ICNN '96*, Washington, 1996.

[Rahmel, 1997] J. Rahmel. *Topology Preserving Neural Networks - Connectionist Learning of Structured Knowledge.* PhD thesis, University of Kaiserslautern, April 1997.

[Ritter *et al.*, 1992] H. Ritter, Th. Martinetz, and K. Schulten. *Neural Computation and Self-Organizing Maps.* Addison Wesley, 2nd edition, 1992.

[Villmann *et al.*, 1996] Th. Villmann, R. Der, M. Herrmann, and Th. Martinetz. Topology preservation in self-organizing feature maps: Exact definition and measurement. *IEEE Transactions on Neural Networks*, 1996. To appear.