Learning Fuzzy Information in a Hybrid Connectionist, Symbolic Model

Steve G. Romaniuk and Lawrence O. Hall

Department of Computer Science and Engineering

University of South Florida

Tampa, FL 33620

e-mail: romaniuk@daffy.csee.usf.edu, hall@waterfall.csee.usf.edu

Abstract

An instance-based learning system is presented. SC-net is a fuzzy hybrid connectionist, symbolic learning system. It remembers some examples and makes groups of examples into exemplars. All real-valued attributes are represented as fuzzy sets. The network representation and learning method is described. To illustrate this approach to learning in fuzzy domains, an example of segmenting magnetic resonance images of the brain is discussed. Clearly, the boundaries between human tissues are ill-defined or fuzzy. Example fuzzy rules for recognition are generated. Segmentations are presented that provide results that radiologists find useful.

1 Introduction

This paper describes the use of a hybrid connectionist, symbolic machine learning system, SC-net [4, 8], to learn rules which allow the discrimination of tissues in magnetic resonance (MR) images of the human brain. Specifically, a 5mm thick slice in one spatial orientation will be used to illustrate SC-net's capabilities. The problem involves identifying tissues of interest which include gray matter, white matter, cerebro-spinal fluid (csf), tumor when it exits, edema and/or necrosis. Essentially, a segmentation of the MR image into tissue regions is the aim of this research. The training data is chosen by a radiological technician who is also familiar with image processing and pattern recognition.

SC-net is an instance-based learning system. It encodes instances or modifications of instances in a connectionist architecture for use in classification after learning. Fuzzy sets are directly represented by groups of cells in the network. Membership functions for any defined fuzzy sets are also learned during the training process with the dynamic plateau

modification feature of SC-net [7].

The rest of this paper will consist of a description of the relevant features of the SC-net learning system, a description of the processing of a MR image slice, the presentation and discussion of the segmentation results obtained with the SC-net system, a discussion of how these results compare with other techniques that have been used [5] and an analysis of the feasibility of the SC-net approach in this domain.

2 The SC-net approach

Each cell in an SC-net network is either a min, max, negation or linear threshold cell. The cell activation formulae are shown in Figure 1. The output structure of the network is set up to collect positive and negative evidence for each output. For an output cell in a classificatory domain, an output of 0 indicates no presence, 0.5 indicates unknown and 1 indicates true. We will show an example of a different us of the output values in the MR image segmentation domain.

SC-net configures its connectionist architecture based upon the training examples presented to it. The learning algorithm responsible for the creation of the network topology is the Recruitment of Cells algorithm (RCA) [4, 7]. RCA is an incremental, instance-based algorithm that requires only a single pass through the training set. Every training instance is individually presented to the network for a single feedforward pass. After the pass has been completed, the actual and the expected activation for every output are compared. Three possible conditions may result from this comparison.

- The example was correctly identified (error is below some epsilon). No modifications are made to the network.
- The example is similar to at least one previously seen and stored instance (error within 5 epsilon). For those output cells that have an activation within 5 epsilon of the expected output, a bias is adjusted to incorporate the new instance.
- The example could not be identified by the network. This results in the recruitment of

CA_i – cell activation for cell C_i.
O_i – output for cell C_i in [0,1].
$O_{i_{positive}}$ and $O_{i_{negative}}$ are the positive and negative collector cells for C_i respectively.
$CW_{i,j}$ – weight for connection between cell C_i and C_j, $CW_{i,j}$ in R.
CB_i – cell bias for cell C_i, CB_i in [-1..+1].

$$CA_i = \begin{cases} min_{j=0,..,i-1,i+1,..n}(O_j * CW_{i,j}) * |CB_i| & C_i \text{ is a min cell} \\ max_{j=0,..,i-1,i+1,..n}(O_j * CW_{i,j}) * |CB_i| & C_i \text{ is a max cell} \\ \left|\sum_{j=0;j\neq i}^{n} O_j * CW_{i,j}\right| * CB_i & C_i \text{ is a ltc cell} \\ 1 - (O_j * CW_{i,j}) & C_i \text{ is a negate cell} \\ O_{i_{positive}} + O_{i_{negative}} - 1/2 & C_i \text{ is either an intermediate} \\ & \text{or final output cell.} \end{cases}$$

$$O_i = max(0, min(1, CA_i))$$

Figure 1: Cell activation formula

a new cell (referred to as an information collector cell, ICC). Appropriate connections from the network inputs to the ICC are created. The ICC cell itself is connected to either the positive (PC) or negative collector (NC) cell. The PC is used to collect positive evidence, whereas the NC accumulates negative evidence. The initial empty network structure for a two input (one output) fuzzy exclusive-or is presented in Figure 2. Note that the uk cell always takes an activation of 0.5. The complete learned network for the fuzzy exclusive-or is shown in Figure 3, where cells c1-c3, c5 are IC cells and n1, n2, c4, and c6 are negation cells.

To improve on the generalization capabilities of the RCA generated SC-net network a form of post training generalization is employed. This method is called the min-drop feature. Whenever a test pattern is presented to the system, which cannot be identified by any of the output cells, the min-drop feature is applied. If a new pattern cannot be recognized by the network, all output cells will be in an inactive state (an unknown response of 0.5 is returned). In this case the min-drop feature is applied to find the nearest corresponding

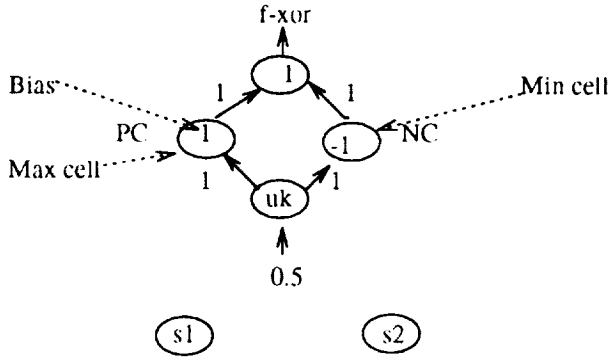


Figure 2: SC-net structure

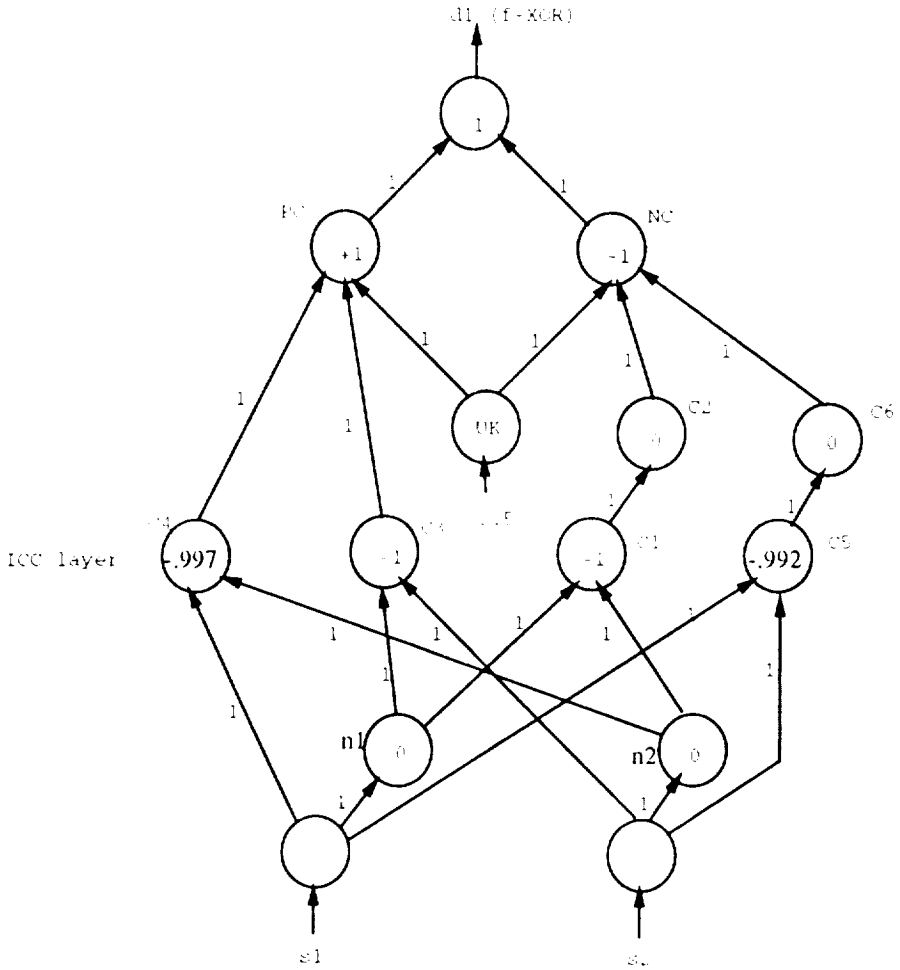


Figure 3: The network for the fuzzy exclusive-or

output for the current pattern. New patterns are stored in the network through recruitment of IC cells (and possibly some negation cells). These IC cells are essentially min-cells, which return the minimum of the product formed from the incoming activation and the weight on the corresponding connection. The min-drop feature works by dropping (ignoring) the next piece of evidence which is below some threshold. The process is repeated until one or more output cells enter an active state (fire). The final number of connections dropped indicates the degree of generalization required to match the newly presented pattern. In a second mode, a bound may be placed on the min-drop value, preventing an unwarranted over-generalization. RCA and post training generalization in the form of the min-drop feature provide good generalization. However, several problems can be associated with the RCA learning phase.

- Network growth can be linear in the number of training examples.
- As a direct consequence of the first problem storage and time (to perform a single feedforward pass) requirements may increase beyond the networks physical limitations.

- Generalization on yet unseen patterns is limited, and requires use of min-drop feature.

To address the above problems a network pruning algorithm was developed. The GAC (Global attribute Covering) algorithm's [7] main purpose is to determine a minimal set of cells and links, which is equivalent to the network generated by RCA. That is, all previously learned information should be retained in the pruned network. GAC attempts to determine a minimal set of connections, which may act as inhibitors of the information collector cells (ICC). Each information collector cell is introduced to the network as the result of an example in the training set which was distinct from all previously seen examples. GAC is completely described in [8].

2.1 Dynamic Plateau Modification of fuzzy membership functions

All fuzzy membership functions in SC-net are represented as trapezoidal fuzzy sets [7, 9]. They are represented in the network by a group of cells as shown in Figure 4 for the fuzzy variable teenager. Teenager takes membership values of 1 in [13..19], of course. In this implementation the membership goes linearly to 0 at the ages of 5 and 25. In the network ages are translated into [0,1] from the [0,100] year range. So the age of 22 is translated to 0.22. Figure 5 shows the actual graph of the membership function for the fuzzy teenager variable.

The dynamic plateau modification function (DPM) is designed to bring in the arms of the fuzzy membership function. In general, we allow the range of the membership function for unknown functions to initially be the range of the fuzzy variable. The range in which the function obtains a value of 1 is at least one point (all fuzzy sets in SC-net are normal in the sense that they contain at least one full member) and usually much smaller than the function range. Hence, for the teenage example with a 100 year range the right arm of the trapezoidal membership function would initially go to 0 at age 100, if we had no information on constructing the membership function other than where it is crisp (attains a membership value of 1). We always assume that the crisp (normal) portion of the membership function is known. The DPM function allows us to arbitrarily set the arms too wide and then adjust them during the learning process. Clearly, in our example it is impractical for someone 99 or 100 years old to have membership in the fuzzy set teenager.

A high-level description of the DPM method is as follows. When it is determined that the fuzzy membership value has caused an incorrect output, the maximal membership that will not cause an error is determined. This value for the set element given and the nearest element at which the membership function takes a value of 1 are used to specify the linear arm of the function. This provides a new upper or lower plateau value (point at which the function goes to 0) for the fuzzy membership function which is used to update the weights

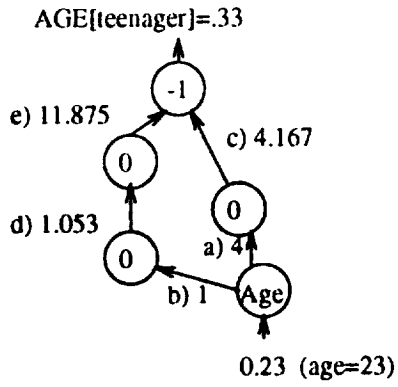


Figure 4: The fuzzy variable teenager.

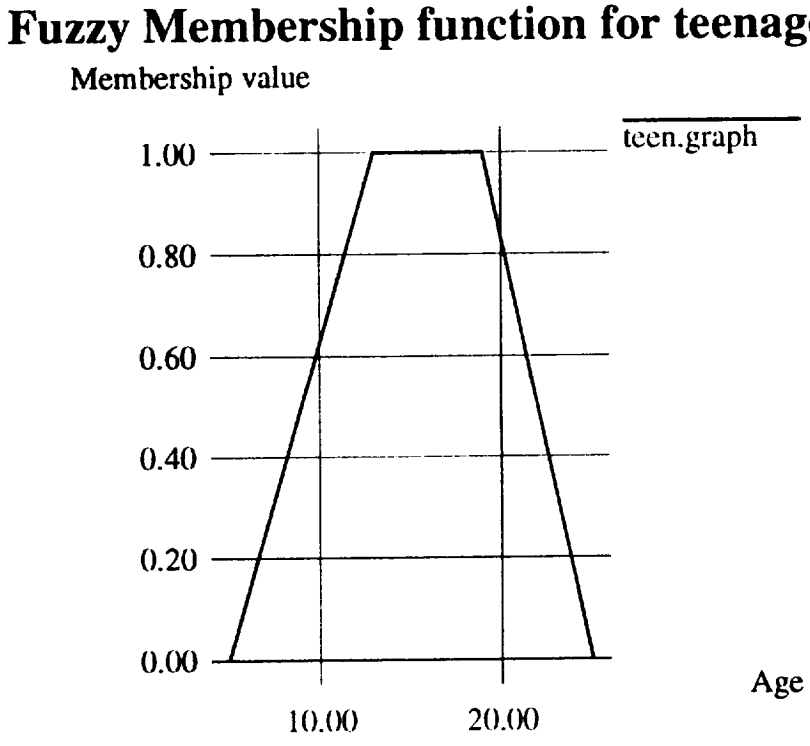


Figure 5: Graph of membership function for fuzzy variable teenager.

labeled a thru e in Figure 4 [9].

2.2 Automatic partition generator

In SC-net all real-valued inputs are modeled by a set of individual fuzzy sets which cover the range of the input. In the case that real-valued data is truly fuzzy, but domain experts do not exist to provide indications of how to model it by fuzzy sets, the choice of the fuzzy sets to cover the range is difficult. Since the data is fuzzy, it may not be possible to accurately identify distinct ranges of the real-valued output associated with specific output. However, this type of idea of associating (fuzzy) ranges with actual outputs can be used. The automatic partition generator (APG) is a method to develop a viable set of fuzzy sets for use in the learning process in domains which have real-valued input, but no expert identified ranges that may belong to specific fuzzy sets.

The APG algorithm works as follows. For each real-valued attribute or feature it makes a partition such that the boundary going from low value to higher value includes at least one element of a class. It will further contain as many elements of the same class as possible. Given the strategy to have all the partitions contain only one class, the maximum number of partitions for any given feature would be the number of classes and would indicate it is very difficult to partition the train set based on that feature or attribute alone. It is the case that a partition may be bounded on both sides by partitions that belong to the same class which is a different class than the examples in the bounded partition belong to.

3 The Nature of MRI Data

Magnetic Resonance Imaging (MRI) systems measure the spatial distribution of several soft tissue related parameters such as T1 relaxation (spin lattice), T2 relaxation (transverse) and proton density. By discrete variations of the radio frequency (RF) timing parameters, a set of images of varying soft tissue contrast can be obtained. The use of time varying magnetic field gradients provide spatial information based on the frequency or phase of the precessing

protons using both multi-slice (2DFT) or volume (3DFT) imaging methods [10, 11]. Hence, a multi-spectral image data set is produced.

In our work, male volunteers (25-45 years) and patient tumor studies were performed on a high field MRI system (1.5 tesla) using a resonator quaduature detector head RF coil. Transverse images of 5 mm thickness were obtained using a standard spin echo (SE) technique for T1 weighted images (pulse repetition time TR = 600 ms, echo time TE=20 ms) and proton density (ρ) and T2 weighted images (TR=3000 ms, TE=20 and 80 ms respectively), using the 2DFT multi-slice technique [12, 13, 2]. Volunteers were imaged for the same anatomical location.

Pixel intensity based classification methods were employed in this work as opposed to methods based on the calculation of magnetic resonance relaxation parameters. The latter methods require tailored RF pulse sequences [10, 11]. Image intensity based methods can be applied to any imaging protocol and are not restricted to the number of images acquired, i.e. it is possible to accommodate images with features other than MR relaxation parameters, such as perfusion and diffusion imaging, metabolic imaging and the addition of images from other diagnostic modalities [2]. The transverse images were acquired, centrally located in the resonator RF head coil, and hence did not require uniformity corrections for RF coil geometry or dielectric loading characteristics as developed at this institute [3]. Similarly, the subjects studied did not move significantly during the imaging procedure and hence, corrections were not required for related registration problems.

4 Segmenting magnetic resonance images

SC-net is a supervised instance-based learning system. Hence, in order to use it to segment an image a training set of labeled pixels must exist. Each pixel has 3 features associated with it a T1, T2 and proton density value. In this paper, we will focus on one normal slice and one abnormal slice. There are 271 pixels in the abnormal training set and 216 pixels in the normal training set. There are 5 classes in the normal train set; gray matter, white

matter, csf, fat and air. The abnormal train set also contains a class for tumor or pathology for a total of 6 classes. Each of the train sets was chosen by a radiological technician.

Each of the input features is real-valued taking values in [0,255] and hence will be represented as fuzzy sets within SC-net. However, it is unclear how these fuzzy sets should be constructed. Further, in [6] it is shown that the values associated with specific tissues vary from subject to subject with significant overlap. Therefore, the partitions of the input ranges for the initial fuzzy sets for each of the inputs were obtained by the use of the APG algorithm.

The inputs in each dimension are first translated from [0,max_value] $max\_value \leq 255$ to the [0,1] range. The APG algorithm is then run which, for example, in the normal (volunteer) training set produces 11 partitions in T1, 19 partitions in proton density (ρ) and 5 partitions in T2. It is interesting that T2 requires the least partitions as it has been the most used single parameter in the literature and few partitions will belong to features or attributes that are "good" data separators. The initial range of each constructed fuzzy set is [-0.2,1.2]. Allowing the range of the membership function to be larger than the range of the set it models is an implementation convention which allows membership values to be 1 at the edges of the actual range.

There are two possible ways to assign examples to classes. One is to use 5 outputs for the normal example and 6 outputs for the abnormal example. This is the most straightforward method. Another possibility exists, which is to use just 1 output. This output is then broken into 5 ranges for the normal example (i.e. [0,0.2], (0.2,0.4], (0.4, 0.6], (0.6, 0.8], and (0.8,1]) which respectively represent the 5 tissue types of interest. Similarly, the single output range can be broken up for 6 outputs. The use of one output provides a very compact network with just 3 inputs which fan out into 35 fuzzy sets in the normal example.

In all experiments, after training all of the remaining pixels are presented to the network for classification. The image is 256 by 256, which means that the training set is very small in relation to the total set of 65,536 pixels.

Table 1: Synthetic Colors for MR Tissue Classes.

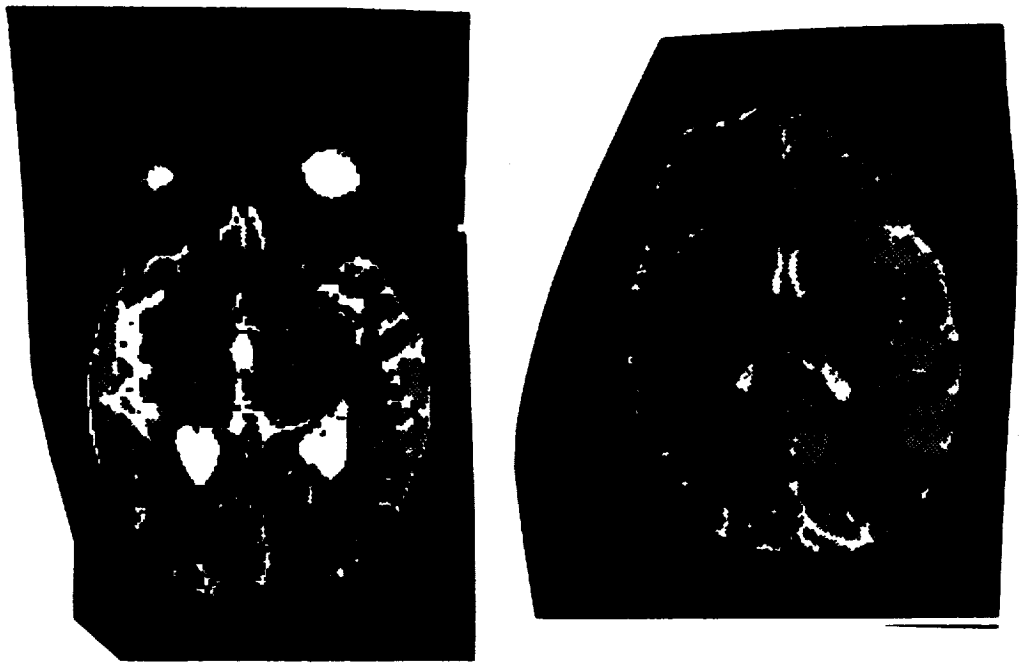
blue	air
yellow	cerebrospinal fluid (csf)
red	white matter
orange	gray matter
brown	fat
purple	pathology

4.1 Results

In Figure 6, we show the segmentation results for a patient with pathology (6a) using 6 outputs and a normal volunteer with 5 outputs (6b). In both cases the fuzzy outputs have been made into one crisp color. The chosen color is the one associated with the output which has the highest membership value. A color table for the figures is listed in Table 1. The patient with pathology has received chemo and radiation therapy which has eliminated obvious tumors, but left some pathology.

The segmentations in Figure 6 are comparable to segmentations pronounced as good by a team of radiologists [5]. The only real difference is that some fat (brown) shows up within the brain. However, this is a minor inconsistency. The case with pathology is segmented as well as any of the other fuzzy unsupervised and non-fuzzy supervised techniques used in [5]. In the lower left-hand part of the image the pathology is clearly defined and it can be seen that there is also pathology in the top of the image and the lower right-hand part of the image.

In Figure 7, we show the results using only 1 output for the abnormal case (7a) and normal case (7b). It can be seen that the segmentations are much the same as before. The fat in 7b is only weakly misclassified in this instance and barely shows up in the segmented image. These displays are fuzzy, which means that a pixel that strongly belongs to a class gets a bright color value, while a pixel that weakly belongs to a class is a darker shade of the same color. This generally shows the uncertainty in the segmentation better and tends



a) b)

Figure 6: An abnormal and normal segmentation by SC-net with multiple outputs.

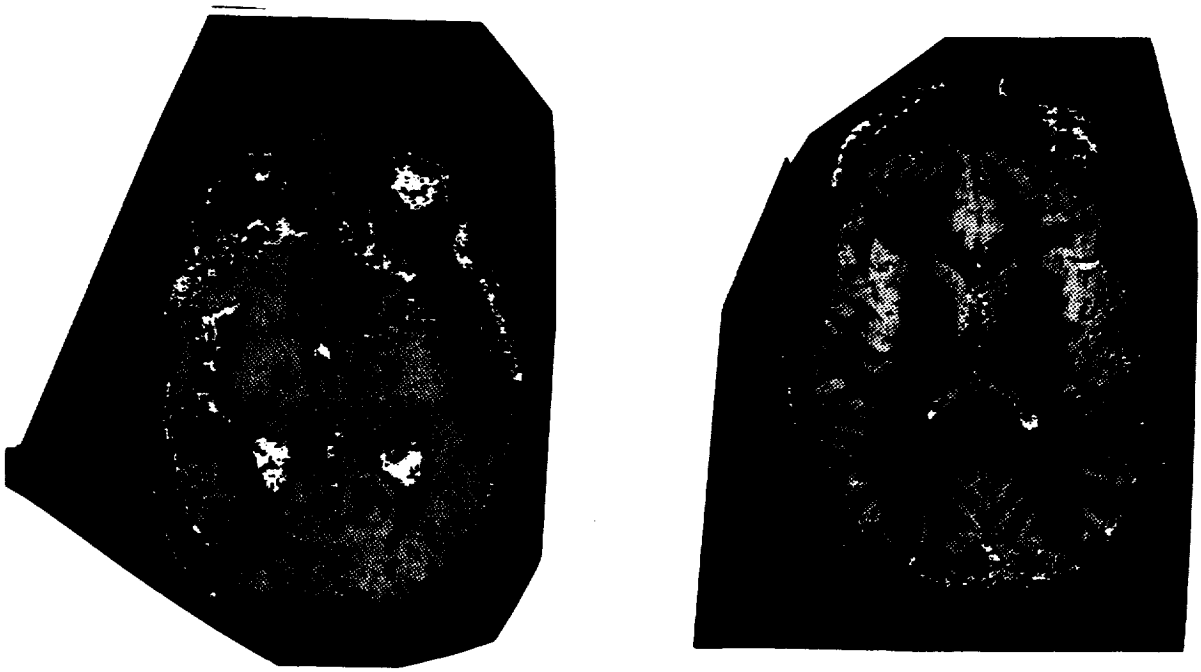
to highlight borders [5].

5 Summary

SC-net is able to provide good segmentations of MR images of the brain. This is a domain in which there is significant tissue overlap and the boundaries are fuzzy. With the use of the APG function the real-valued inputs are automatically partitioned into fuzzy sets. These fuzzy sets are further refined after the RCA learning algorithm has been applied by the use of DPM.

The results of the segmentation are comparable to those obtained by K nearest neighbor (K-nn) (K=7) and Cascade Correlation [5] in another study of supervised learning techniques. In the normal volunteer image the SC-net segmentation is a little clearer than the k-nn segmentation with the one exception of misclassified fat. The fuzzy connectionist representation of SC-net is very effective and fast in learning and classifying the MR images. The rules that are generated after the use of GAC for the normal case numbered 9 and 13 for the abnormal case. They can be used to provide a sense of what portions of which features are important in the recognition process. In Figure 8, the 9 rules for a normal case are shown. It can be seen that for output 5, fat, the 16^{th} partition of the T2 parameter is crucial. For output 2, csf, around the 2^{nd} proton density partition is the an important indicator. Output 1, which is air, is very easy to distinguish by one rule. This is a known fact since it essentially has a 0 return. The number of rules required to distinguish a class can also be an indication of how difficult it is to recognize. Hence, the rules can have semantic meaning and may be useful in tuning the system which is an advantage of a hybrid representation.

Acknowledgements: Thanks to Robert Velthuizen for helping us with the image display and providing an expert interpretation of the images. This research was partially supported by a grant from the Whitaker Foundation.



a)

b)

Figure 7: An abnormal and normal segmentation by SC-net with 1 output.

```
Rule 1:  if and( fuzzy(I3[p16]) = 1.000, fuzzy(I2[p12]) = 1.000,
fuzzy(I1[p5]) = 1.000 ) then Out5 ( 1.000 ).

Rule 2:  if and( fuzzy(I3[p16]) = 1.000, fuzzy(I2[p17]) = 1.000,
fuzzy(I1[p7]) = 1.000 ) then Out5 ( 1.000 ).

Rule 3:  if and( fuzzy(I3[p16]) = 1.000, fuzzy(I2[p19]) = 1.000,
fuzzy(I1[p16]) =  1.000 ) then Out5 ( 1.000 ).

Rule 4:  if and( fuzzy(I3[p2]) = 1.000, fuzzy(I2[p2]) = 1.000,
fuzzy(I1[p17]) = 1.000 ) then Out4 ( 1.000 ).

Rule 5:  if and( fuzzy(I3[p15]) = 1.000, fuzzy(I2[p3]) = 1.000,
fuzzy(I1[p17]) = 1.000 ) then Out3 ( 1.000 ).

Rule 6:  if and( fuzzy(I3[p22]) = 1.000, fuzzy(I2[p3]) = 1.000,
fuzzy(I1[p5]) = 1.000 ) then Out2 ( 1.000 ).

Rule 7:  if and( fuzzy(I3[p17]) = 1.000, fuzzy(I2[p2]) = 1.000,
fuzzy(I1[p5]) = 1.000 ) then Out2 ( 1.000 ).

Rule 8:  if and( fuzzy(I3[p19]) = 1.000, fuzzy(I2[p2]) = 1.000,
fuzzy(I1[p6]) = 1.000 ) then Out2 ( 1.000 ).

Rule 9:  if and( fuzzy(I3[p1]) = 1.000, fuzzy(I2[p1]) = 1.000,
fuzzy(I1[p1]) = 1.000 ) then Out1 ( 1.000 ).
```

Figure 8: Rules for normal volunteer

References

[1] Bensaid, A. M., Hall, L. O., Clarke, L. P., and Velthuizen, R. P. MRI Segmentation Using Supervised and Unsupervised Methods. IEEE Eng. Med. & Biol., Proceedings of 13th Annual Meeting, Oct. 31st-Nov. 3rd, Orlando, Florida, 1991.

[2] Clarke, L. P. Comparison of Bayesian Maximum Likelihood (MLM) and Artificial Neural Network for Supervised Tissue Classification and 3D Segmentation by MRI. Scientific Symposium, American Association of Physicists in Medicine (AAPM), 33rd Annual Meeting, San Francisco, California, July 1991; Med. Physics, 18(3), 673, 1991.

[3] Glennon, D. T., Clarke, L. P., Velthuizen, R. P. and Silbiger, M. L. MR Image Non-Uniformity Correction Techniques. IEEE Eng. Med. Biol., Proceedings of 13th Annual Meeting, Oct. 31st-Nov. 3rd, Orlando, Florida, 1991.

[4] Hall, L.O. and Romaniuk, S.G., " A hybrid connectionist, symbolic learning system", *AAAI-90*, Boston, Ma. 1990.

[5] Hall, L.O., Bensaid, A., Bezdek, J.C., Clarke, L., Silbiger, M. and Velthuizen, R. *A Comparison of Neural Network and Fuzzy Clustering Techniques in Segmenting Magnetic Resonance Images of the Brain*, IEEE Transactions on Neural Networks, V. 3, No. 5, 1992.

[6] Li, C., Goldgof, D. and Hall, L.O., *Towards Automatic Classification and Tissue Labeling of MR Brain Images*, International Association for Pattern Recognition Workshop on Structural and Syntactic Pattern Recognition, Bern, Switzerland.

[7] Romaniuk, S.G. " Extracting Knowledge from a Hybrid Symbolic, Connectionist Network", Ph.D. Dissertation, Department of Computer Science and Engineering, University of South Florida, 1991.

[8] Romaniuk, S.G. and Hall, L.O. "SC-net: A Hybrid Connectionist, Symbolic Network", *Information Sciences*, To Appear, 1993.

[9] Romaniuk, S.G. and Hall, L.O. "Learning Fuzzy Information in a Hybrid Connectionist, Symbolic Model", *IEEE International Conference on Fuzzy Systems 1992*, pp. 309-312., San Diego, Ca. 1992.

[10] Bottomley, P. A., Foster, T. H., Argersinger, R. E. and Pfeifer, L. M. A. Review of Normal Tissue Hydrogen NMR Relaxation Times and Relaxation Mechanisms from 1-100 MHz: Dependence on Tissue Type, NMR Frequency, Temperature, Species, excision and Age. Medical Physics, 11(4), 425-448, 1984.

[11] Hyman, T. J., Kurland, R. J., Leroy, G. C. and Shoop, J. Characterization of Normal Brain Tissue Using Seven Calculated MRI Parameters and a Statistical Analysis System. Mag. Res. Med. 11, 22-34 1989.

[12] Bensaid, A. M., Hall, L. O., Clarke, L. P., and Velthuizen, R. P. MRI Segmentation Using Supervised and Unsupervised Methods. IEEE Eng. Med. & Biol., Proceedings of 13th Annual Meeting, Oct. 31st-Nov. 3rd, Orlando, Florida, 1991.

[13] Schellenberg, J. D., Naylor, W. C. and Clarke, L. P. Application of Artificial Neural Networks for Tissue Classification from Multispectral Magnetic Resonance Images of the Head. IEEE Symposium, Computer-Based Medical Systems, Third Annual Conference Proceedings, Chapel Hill, North Carolina, June 3-6, 1990.