

A Hybrid Approach to MR Imaging Segmentation Using Unsupervised Clustering and Approximate Reducts

Sebastian Widz¹, Kenneth Revett², Dominik Ślęzak³

¹ Deforma Sebastian Widz, Warsaw, Poland

² University of Westminster, London, UK

³ University of Regina, Regina, Canada

Abstract. We introduce a hybrid approach to magnetic resonance image segmentation using unsupervised clustering and the rules derived from approximate decision reducts. We utilize the MRI phantoms from the Simulated Brain Database. We run experiments on randomly selected slices from a volumetric set of multi-modal MR images (T1, T2, PD). Segmentation accuracy reaches 96% for the highest resolution images and 89% for the noisiest image volume. We also tested the resultant classifier on real clinical data, which yielded an accuracy of approximately 84%.

Keywords: MRI segmentation, rough sets, approximate decision reducts, self-organizing maps, magnitude histogram clustering.

1 Introduction

Segmentation is the process of assigning the class labels to data containing spatially varying information. For Magnetic Resonance Imaging (MRI), we have a 3D data set (a volume) collected as a series of 2D slices. The goal is to classify every voxel within a slice to one of the tissue classes. We focus on three classes of a clinical interest: cerebrospinal fluid (CSF), grey matter (GM), and white matter (WM). It has been demonstrated repeatedly in the literature that relative distribution of these classes is diagnostic for specific diseases such as stroke, Alzheimer’s disease, various forms of dementia, and multiple sclerosis [6, 7]. An automated method for segmenting tissue would provide a useful adjunct to clinical radiology for the effective diagnosis of disease.

Rough sets [1, 9, 14] provide powerful methodology for constructing classification models, useful also for image segmentation (cf. [5]). In our approach, we create a decision table with objects corresponding to voxels and decisions derived from the Simulated Brain Database (SBD) of fully segmented images¹ [2, 12]. To define conditional attributes, we employ two unsupervised clustering algorithms – the self-organizing map and our own, simple method for the magnitude histogram clustering. Using an order-based genetic algorithm (o-GA) [4, 22], we search for the optimal R -approximate reducts – irreducible subsets of

¹ Provided by Montreal Neurological Institute <http://www.bic.mni.mcgill.ca/brainweb>

conditional attributes, which approximately preserve the data-based global relative gain of decision [17, 18]. From the best found reducts we generate (possibly inexact) "if...then..." rules applicable to classification of voxels in new slices.

There are several imaging parameters that affect resolution of MRI: slice thickness, noise levels, bias field inhomogeneities (INU), partial volume effects, imaging modality. These factors should be incorporated into an automated segmentation system in order for it be clinically relevant. Hence, we verify the usage of reducts learned over the slices from a high resolution image volume (1mm, 3% noise, 20% INU) also for images with possibly higher noise (up to 9%) and INU (up to 40%). Comparing to our previous research [19, 20], where the obtained decision rules were tested only against the SBD data, we applied our segmentation algorithm on real clinical images (T2-weighted, 5mm thick slices with relatively low noise and bias). Despite differences in characteristics of the training and testing images, we obtained a segmentation accuracy of approximately 84% on clinical images. We believe this is a result of both appropriately derived attributes and robustness of short rules induced from approximate reducts.

The paper is organized as follows: In Section 2, we describe the basic data preparation techniques. In Section 3, we introduce the details of the applied unsupervised clustering algorithms. In Section 4 we discuss the foundations of the applied approximate attribute reduction method. In Section 4, we present the experimental results. In Section 5 we summarize our work.

2 Data Preparation

The primary source of information obtainable from an MR Image is the magnitude value for a given voxel. For reasonably good MR images, for every considered modality, the voxels form a series of Gaussian distributions corresponding to the tissue classes. In a typical MR image of the brain, there are generally the following classes: bone, Cerebral Spinal Fluid (CSF), Grey Matter (GM), White Matter (WM), as well as fat, skin, muscle and background. An example of magnitude histogram is illustrated in Fig. 1. In this study, we focus on CSF, GM, and WM exclusively. Therefore, the first processing step was to remove all voxels that do not belong to the actual brain. Background is removed automatically by applying a simple mask algorithm. Bone, fat, skin and muscles are removed manually, basing on information about the SBD image crisp phantom.

To apply the segmentation procedure based on decision rules derived from data, we must first generate the training data set. Following the standards developed within the theory of rough sets [14], we form a decision table $\mathbb{A} = (U, A \cup \{d\})$, where each attribute $a \in A \cup \{d\}$ is a function $a : U \rightarrow V_a$ from the universe U into the value set V_a . The elements of U are voxels taken from MR images. Each SBD data set is a series of 181 217x181 images. We use a range $\langle 61;130 \rangle$ of slices where we have good representation of all tissue classes. The decision attribute $d \notin A$ represents the ground truth, which is necessary during the classifier training phase. The set A should contain the attributes labeling the voxels with respect to the available MRI modalities. Further, we characterize the method we employed to extract the attributes in A from the MRI images.

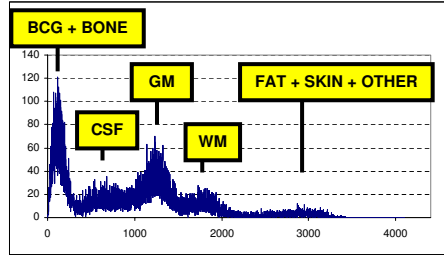


Fig. 1. A single bin frequency histogram from a T1 SBD slice #81 (1mm slice thickness, 3% noise and 20% INU). The x-axis values are 12 bit-unsigned integers, corresponding to the magnitude of the voxels from the raw image data. The histogram’s peaks are likely to correspond to particular decision/tissue classes.

Edge attributes are denoted by $edgeT1$, $edgeT2$, $edgePD$. They are derived using a simple *Discrete Laplacian* method – a general non-directional gradient operator determining whether the neighborhood of a voxel is homogenous. For instance, $edgeT1$ takes the value 0 for a given voxel, if its neighborhood for T1 is homogeneous, and 1 otherwise. We apply the 3x3 window to determine the voxel neighbors. We use a threshold determined by the noise level parameter estimated by running various tests on T1, T2 and PD training images, respectively. For each parameter value at the range [0,0.5] we run a test measuring correctness of recognized edges versus the phantom edges. Generally, the optimal noise parameter setting for T1 is 0.08, while for T2 and PD it is 0.13. The edge attributes provide information that may be subject to a *partial volume effect* (PVE) – It can be used also in the future studies on PVE.

Magnitude attributes are denoted by $hcMagT1$, $hcMagT2$, $hcMagPD$, as well as $somMagT1$, $somMagT2$, $somMagPD$. They are derived using the *histogram clustering algorithm* HCLUSTER (prefix ‘*hc*’) and self-organizing map SOM (prefix ‘*som*’) for all three image modalities. SOM and HCLUSTER perform the unsupervised segmentation of the image – they are discussed later in Section 3. The results of such segmentation are recorded as the corresponding attribute values. In our previous work [19, 20], we employed a manually based approach using polynomial interpolation to find the *Full-Width Half-Maximum* (FWHM) value for each of the histogram peaks (cf. [2]). We find HCLUSTER and SOM techniques more autonomous and accurate than FWHM.

Neighbor attributes are denoted by $hcNbrT1$, $hcNbrT2$, $hcNbrPD$, as well as $somNbrT1$, $somNbrT2$, $somNbrPD$. They are derived as follows:

```

If (EDGE == 0)
    NBR = the same class as in case of the MAG attribute
Else // EDGE == 1
    NBR = major class value appearing in the neighborhood

```

Naturally, for *hc*-attributes we use the class values derived from the HCLUSTER algorithm, and for *som*-attributes – those provided by SOM. For example, the value of $hcNbrT1$ is calculated from $edgeT1$ (EDGE) and $hcMagT1$ (MAG).

Mask attribute is denoted as *msk*. It is used to mask the brain area. It can happen that two different tissues have similar voxel value and appear on the histogram in one peak. Knowing the relative position of a voxel may help in removing the resulting classification ambiguity. It provides us with extra information about the voxel location (e.g. CSF tissue is located more in the middle of the brain). For the exact algorithm of deriving the mask values we refer to our previous work [19, 20]. Here, we illustrate its appearance in Fig. 2D.

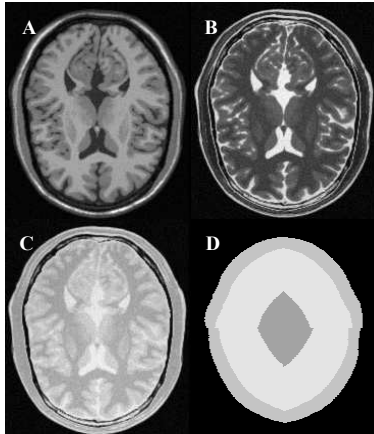


Fig. 2. Modalities T1 (Picture A), T2 (B), and PD (C) from the SBD slice #81 (thickness 1mm, 3% noise, 20% INU). Picture D presents the mask obtained for this case.

3 Histogram Clustering and Self-Organizing Maps

We propose a simple method for clustering the magnitude histograms, referred as HCLUSTER. Let us denote the value of the i -th histogram's position by h_i . We define the distance $dist(i, j)$ between positions i and j as the length of the path between points (i, h_i) and (j, h_j) with the histogram. Let's assume that $i < j$. We have $dist(i, j) = \sum_{k=i}^{j-1} \sqrt{1 + (h_k - h_{k+1})^2}$, as illustrated by Figure 3. The algorithm is controlled by parameters α and β , which influence the number of found clusters, as well as sensitivity to the histogram's peaks.

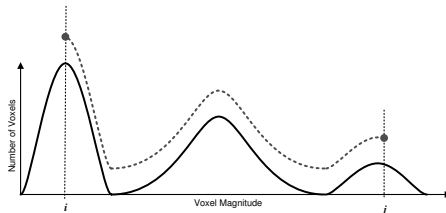


Fig. 3. Illustration of histogram-based distance $dist(i, j)$ between positions i and j .

(α, β) -HCLUSTER

1. Let DOM be the histogram's domain, REP – the set of the cluster representatives, and CAN – the set of candidates. Put $CAN = DOM$, $REP = \emptyset$.
2. Add position of the highest peak to REP , as the first cluster representative.
3. *while* ($\max_{i \in CAN} (\min_{j \in REP} dist(i, j) + \alpha * h_i) \geq \beta$)
 $i_{rep} = argmax_{i \in CAN} (\min_{j \in REP} dist(i, j) + \alpha * h_i)$
 $CAN = CAN \setminus \{i_{rep}\}$
 $REP = REP \cup \{i_{rep}\}$
4. Put every element of DOM to the cluster corresponding to its closest (due to distance $dist$) element of REP .

For every modality, the obtained cluster numbers correspond to the magnitude attribute values in the decision table. For instance, if the PD -magnitude of the given voxel $u \in U$ drops into the 3rd cluster (counting from left hand side of the PD 's histogram), then we get the value $hcMagPD(u) = 2$.

The self-organizing map (SOM) is employed to automate the association of a class label based on the magnitude of a voxel in a similar way as HCLUSTER. The network consists of 12 nodes, each initialized with normalized random weights. The inputs are normalized magnitude values of voxels extracted from images. The whole framework may work with either single or multiple modalities, although, in this paper, we apply it only separately to T1, T2 and PD.

The SOM is trained on a single row in the middle of the randomly selected slice from SBD. The weights are updated using the least mean square algorithm. The training is terminated when the error is below a user specified level (0.01). After training, the network is used to classify and generate values for $somMag$ and $somNbr$ attributes for all voxels within the given volume.

Due to space constraints, we do not discuss how to adjust parameters of the self-organizing map (like neighborhood, learning rate, maximum number of learning cycles, etc.) and HCLUSTER (α and β). In the same way, we will not discuss parameters of the genetic algorithm described in the next section. All these settings are adjusted due to our experience with data processing.

4 Approximate Reduction

When modeling complex phenomena, one must strike a balance between accuracy and computational complexity. This balance can be achieved through the use of a *decision reduct*: an irreducible subset $B \subseteq A$ determining d in decision table $\mathbb{A} = (U, A \cup \{d\})$. The obtained decision reducts are used to produce the decision rules from the training data. Reducts generating smaller number of rules seem to be better as the corresponding rules are more general and applicable.

Sometimes it is better to remove more attributes to get even shorter rules at the cost of their slight inconsistencies. One can specify an arbitrary data-based measure M which evaluates a degree of influence $M(d/B)$ of subsets $B \subseteq A$ on d . Then one can decide which attributes may be removed from A without a *significant* loss of M . Given the approximation threshold $\varepsilon \in [0, 1)$, let us say

that $B \subseteq A$ is an (M, ε) -approximate decision reduct, if and only if it satisfies inequality $M(d/B) \geq (1 - \varepsilon) * M(d/A)$ and none of its proper subsets does it. For an advanced study on such reducts we refer the reader to [15, 16]. Here we consider the *multi-decision relative gain measure* [17, 18]:

$$R(d/B) = \sum_{\text{rules } r \text{ induced by } B} \left(\frac{\text{number of objects recognizable by } r}{\text{number of objects in } U} * \max_i \frac{\text{probability of the } i\text{-th decision class induced by } r}{\text{prior probability of the } i\text{-th decision class}} \right) \quad (1)$$

Measure (1) expresses the average gain in determining decision classes under the evidence provided by the rules generated by $B \subseteq A$. It can be used, e.g., to evaluate the potential influence of a particular attributes on the decision. The quantities of $R(d/\{a\})$, $a \in A$, reflect the average information gain obtained from one-attribute rules. They are, however, not enough to select the *subsets* of relevant attributes. For instance, several attributes $a \in A$ with low values of $R(d/\{a\})$ can create together a subset $B \subseteq A$ with high $R(d/B)$.

The problems of finding approximate reducts are generally hard (cf. [15]). Therefore, for the decision table with attributes described in the previous sections, we prefer to consider the use of a heuristic rather than an exhaustive search. We adapt the *order based genetic algorithm (o-GA)* for searching for minimal decision reducts [22] to find heuristically (sub)optimal (R, ε) -approximate decision reducts. We follow the same way of extension as that proposed in [16], also used by ourselves in the previous papers on MRI segmentation [19, 20]. As a *hybrid algorithm* [4], the applied o-GA consists of two parts:

1. *Genetic part*, where each chromosome encodes a permutation of attributes
2. *Heuristic part*, where permutations τ are put into the following algorithm:

(R, ε) -REDORD algorithm (cf. [16, 22])

1. Let $\mathbb{A} = (U, A \cup \{d\})$ and $\tau : \{1, \dots, |A|\} \rightarrow \{1, \dots, |A|\}$ be given; Let $B_\tau = A$;
2. For $i = 1$ to $|A|$ repeat steps 3 and 4;
3. Let $B_\tau \leftarrow B_\tau \setminus \{a_{\tau(i)}\}$;
4. If B_τ does not satisfy inequality $R(d/B) \geq (1 - \varepsilon) * R(d/A)$, undo step 3.

We define fitness of a given permutation-individual τ due to the quality of B_τ resulting from (R, ε) -REDORD. As mentioned before, we would like to focus on reducts inducing possibly low numbers of rules. Therefore, we use the following:

$$fitness(\tau) = 1 - \frac{\text{number of rules derived from } B_\tau}{\text{number of rules derived from the whole set } A} \quad (2)$$

To work on permutations-individuals, we use the order cross-over (OX) and the standard mutation switching randomly selected genes [4, 13]. It is important that the results are *always* (R, ε) -approximate decision reducts.

5 Results of Experiments

All the classification tests were performed using the following procedure:

1. Generate the approximate reducts using o-GA based on (R, ϵ) -REDORD for a given $\epsilon \in [0, 1)$. Leave out ten reducts related to the best fitness (2).
2. For every reduct (out of ten left) generate (possibly inexact) decision rules. Filter the rules according to their coverage measure.
3. For every new voxel, vote using the confidence of all applicable rules.

Reducts and rules were generated using 10 slices chosen randomly from the SBD database, for the slice range [61, 130], 1mm thickness, 3% noise, and 20% INU. The resulting classifier were tested against 20 slices (always different than those chosen for training) from the same range, taken from the following volumes:

- Volume 1: 1mm thickness, 3% noise, 20% INU
- Volume 2: 1mm thickness, 9% noise, 20% INU
- Volume 3: 1mm thickness, 9% noise, 40% INU

Obviously, the segmentation is more challenging when applied to slices with a higher noise/INU. Figure 4 presents the average (for 10 experiments repeated with different training/testing slices) accuracy (the number of correctly recognized voxels divided by the number of total number of voxels) and coverage (the number of recognized voxels divided by the total number of voxels).

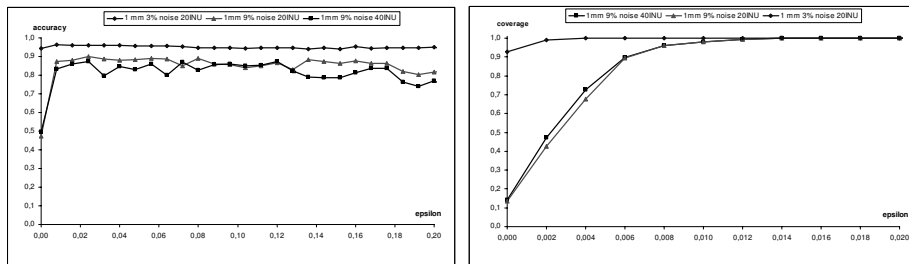


Fig. 4. Accuracy and coverage for the SBD volumes 1,2,3, given different values of ϵ .

For this particular data, given $\epsilon = 0$, there is only one reduct – the set of all attributes. The corresponding rules are then too specific and, therefore, coverage is very low. While increasing ϵ , the reducts become shorter and the number of generated rules decreases, as shown in Figure 5. Coverage grows fast to 1.00 (note that in Figure 4 the range for coverage is 0.00-0.02, while the range for accuracy is 0.0-0.2). The best accuracy is obtained for the following settings:

- Volume 1: 96% for $\epsilon = 0.010$
- Volume 2: 91% for $\epsilon = 0.052$
- Volume 3: 88% for $\epsilon = 0.108$

Let us note that the above results are significantly better than those obtained in [19, 20] (95.1%, 84.4%, 78.7%) using the FWHM-based attributes. Also notice that when the noise and INU increase, higher ϵ gives better performance.

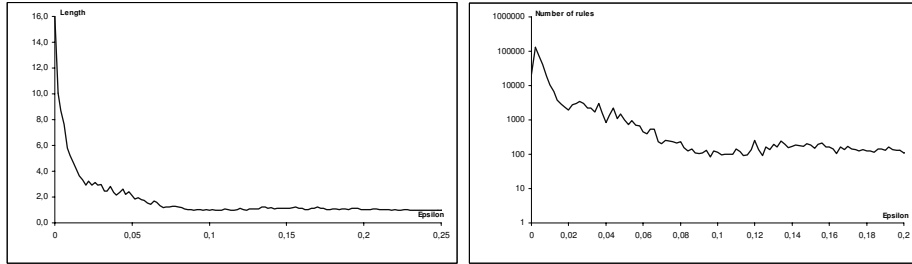


Fig. 5. Average reduct length and number of generated rules for different values of ϵ .

We tested the segmentation accuracy also on a sample of T2-weighted images acquired from a clinical environment. The images were acquired using a transmit and receive birdcage head coil imaged in the axial plane, using an interleaved spin-echo pulse sequence (TR = 130 ms and TE = 3000 ms), 96x96 matrix, slice thickness of 5mm with 1mm inter-slice gap, field of view = 24x24cm with an in-plane resolution of 3.24mm² (1.8 x 1.8mm). No preprocessing of the images was performed. Manual segmentation was performed by a domain expert. Differences between the classifier based segmented tissue classes and clinical tissue classes are reported in Fig. 6. Segmented images are illustrated in Fig. 7.

Tissue Category	Manual Segm.		Autom. Segm.	
	Image 1	Image 2	Image 1	Image 2
Whole Brain	5080	5058	5114	5170
CSF	280	541	200 (1.6%)	403 (2.8%)
GM	2382	2457	2801 (8.2%)	2777 (6.3%)
WM	2409	1707	2113 (5.8%)	2090 (7.6%)
Total Error			15.5%	16.7%

Fig. 6. Comparison between manual and automated segmentation on the T2-weighted axial images. The values in the segmentation columns refers to the total number of voxels n that class and the values in parentheses represents the difference between the manual and automated classification (the associated error) expressed as a percentage of the total number of manually segmented (clinical) voxels.

6 Conclusions and Discussion

The segmentation algorithm developed in this paper was tested on the SBD data sets, producing a maximal accuracy on the order of 96%. We also tested the accuracy of the rule set generated from training on the SBD volumes on real clinically derived T2-weighted images, producing the average accuracy ca. 84%. The primary factor influencing the classification accuracy of the clinical images appears to be directly related to partial volume effect (PVE) – as most voxels that were misclassified were edge voxels – principally between GM and CSF and also between GM and WM. These edge effects are difficult to manually classify – as the domain expert also will face difficulties in deciding, without any preprocessing of the images – which category a voxel belongs to.

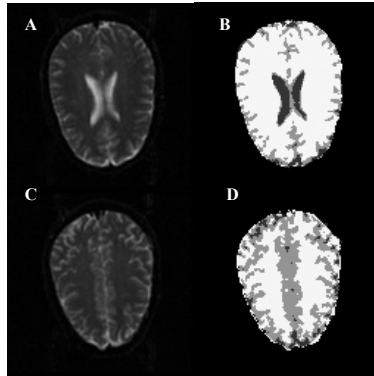


Fig. 7. Real T2-weighted images (A,C) and corresponding segmentation results (B,D).

A favorable factor is that any errors produced by PVE will be systematic – and with respect to comparing segmentation results – one would expect the bias to occur consistently between manual and artificially segmented images. In order to reduce this burden on segmentation accuracy, and to improve the accuracy of our automated segmentation algorithm, we will employ methods that reduce the partial volume effect in our future work. We also plan to train the system on more different data sets. We think that doing that and tuning the parameters better would significantly improve the results.

Our automated MR image segmentation was substantially improved comparing to previous version [19, 20]. A principal change is the automation of the magnitude attribute extraction, previously performed manually through the use of a polynomial interpolation technique. We implemented two automated unsupervised clustering techniques: one based on a derivation of histogram clustering and the other based on a self-organizing map. Because of the lack of space, we did not present the results related to relative importance of the obtained attributes in reducts and rules. Our general observation is that the best classification performance is provided by diversified ensembles of the attribute sets, generated from both applied clustering methods in the same time.

Acknowledgments: The third author was partially supported by the research grant of the Natural Sciences and Engineering Research Council of Canada.

References

1. Bazan, J.G., Nguyen, H.S., Nguyen, S.H., Synak, P., Wróblewski, J.: Rough Set Algorithms in Classification Problem. In: Polkowski, L., Tsumoto, S., Lin, T.Y. (eds): *Rough Set Methods and Applications: New Developments in Knowledge Discovery in Information Systems*. Physica Verlag (2000) pp. 49–88.
2. Cocosco, C.A., Zijdenbos, A.P., Evans, A.C.: Automatic Generation of Training Data for Brain Tissue Classification from MRI. In: *Proc. of MICCAI'2002* (2002).
3. Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, C.J., Evans, A.C.: Design and Construction of a Realistic Digital Brain Phantom *IEEE Transactions on Medical Imaging* 17(3) (1998) pp. 463–468.

4. Davis, L. (ed.): Handbook of Genetic Algorithms. Van Nostrand Reinhold (1991).
5. Hirano S., Tsumoto S.: Segmentation of Medical Images Based on Approximations in Rough Set Theory. In: Proc. of RSCTC'2002. Malvern, USA (2002).
6. Kamber, M., Shinghal, R., Collins, L.: Model-based 3D Segmentation of Multiple Sclerosis Lesions in Magnetic Resonance Brain Images. IEEE Trans Med Imaging 14(3) (1995) pp. 442–453.
7. Kaus, M., Warfield, S.K., Nabavi, A., Black, P.M., Jolesz, F.A., Kikinis, R.: Automated Segmentation of MRI of Brain Tumors. Radiology 218 (2001) pp. 586–591.
8. Kollokian, V.: Performance Analysis of Automatic Techniques for Tissue Classification in Magnetic Resonance Images of the Human Brain. Master's thesis, Concordia University, Montreal, Canada (1996).
9. Komorowski, J., Pawlak, Z., Polkowski, L., Skowron, A.: Rough sets: A tutorial. In: S.K. Pal, A. Skowron (eds): Rough Fuzzy Hybridization – A New Trend in Decision Making. Springer Verlag (1999) pp. 3–98.
10. Kovacevic, N., Lobaugh, N.J., Bronskill, M.J., Levine, B., Feinstein, A., Black, S.E.: A Robust Extraction and Automatic Segmentation of Brain Images. NeuroImage 17 (2002) pp. 1087–1100.
11. Kwan, R.K.S., Evans, A.C., Pike, G.B.: An Extensible MRI Simulator for Post-Processing Evaluation. Visualization in Biomedical Computing (VBC'96). Lecture Notes in Computer Science, vol. 1131. Springer-Verlag, (1996) pp. 135–140
12. Kwan, R.K.S., Evans, A.C., Pike, G.B.: MRI Simulation-Based Evaluation of Image-Processing and Classification Methods. Neuroimage 10 (1999) pp. 417–429.
13. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. Springer-Verlag (1994).
14. Pawlak, Z.: Rough sets – Theoretical aspects of reasoning about data. Kluwer (1991).
15. Ślęzak, D.: Approximate Entropy Reducts. Fundamenta Informaticae (2002).
16. Ślęzak, D., Wróblewski, J.: Order-based genetic algorithms for the search of approximate entropy reducts. In: Proc. of RSFDGrC'2003. Chongqing, China (2003).
17. Ślęzak, D., Ziarko, W.: Attribute Reduction in Bayesian Version of Variable Precision Rough Set Model. In: Proc. of RSKD'2003. Elsevier, ENTCS 82(4) (2003).
18. Ślęzak, D., Ziarko, W.: The investigation of the Bayesian rough set model. Int. J. of Approximate Reasoning, in press.
19. Widz, S., Revett, K., Ślęzak, D.: Application of Rough Set Based Dynamic Parameter Optimization to MRI Segmentation. In: Proc. of 23rd International Conference of the North American Fuzzy Information Processing Society (NAFIPS'2004). Banff, Canada, June 27-30 (2004).
20. Widz, S., Ślęzak, D., Revett, K.: An Automated Multispectral MRI Segmentation Algorithm Using Approximate Reducts. In: Proc. of the 4th International Conference on Rough Sets and Current Trends in Computing (RSCTC'2004). Uppsala, Sweden, June 1-5. Springer Verlag, LNAI 3066 (2004) pp. 815–824.
21. Vannier, M.W.: Validation of Magnetic Resonance Imaging (MRI) Multispectral Tissue Classification. Computerized Medical Imaging and Graphics 15(4) (1991) pp. 217–223.
22. Wróblewski, J.: Theoretical Foundations of Order-Based Genetic Algorithms. Fundamenta Informaticae 28(3-4) (1996) pp. 423–430.
23. Xue, J.H., Pizurica, A., Philips, W., Kerre, E., Van de Walle, R., Lemahieu, I.: An Integrated Method of Adaptive Enhancement for Unsupervised Segmentation of MRI Brain Images. Pattern Recognition Letters 24(15) (2003) pp. 2549–2560.