

Granular Attribute Selection: A Case Study of Rough Set Approach to MRI Segmentation

Sebastian Widz¹ and Dominik Ślęzak²

¹ Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland

² Institute of Mathematics, University of Warsaw
ul. Banacha 2, 02-097 Warsaw, Poland

sebastian.widz@ibspan.waw.pl, slezak@mimuw.edu.pl

Abstract. We introduce a new rough set inspired approach to attribute selection. We consider decision systems with attributes specified by means of two layers: 1) general meta-attribute descriptions, and 2) their specific realizations obtained by setting up parameters of procedures calculating attribute values. We adopt methods designed for finding rough set reducts within the sets of attributes grouped into clusters, where each cluster contains potentially infinite amount of attributes realizing a single meta-attribute. As a case study, we discuss a rough set framework for multi-spectral Magnetic Resonance Image (MRI) segmentation.

Keywords: Rough set reducts, Attribute hierarchies, MRI segmentation.

1 Introduction

Attribute selection is an important step in the process of constructing knowledge representation and classification models [1]. In rough sets, this step corresponds to excluding irrelevant attributes and, as a result, simplifying descriptions of objects in information systems and decision tables [2]. There are numerous approaches to searching for rough set reducts – the sets of attributes containing no irrelevant elements [3]. Rough set attribute selection and classification techniques can be also embedded into hierarchical systems that utilize domain knowledge to approximate complex concepts [4].

In [5], the clustering methods are used to identify groups of attributes that can replace each other in reducts. By working with such attribute granules, one can make the reduct derivation process faster and more intuitive. In this paper, we note that attributes can be grouped also with respect to other criteria. Quite often, domain experts sketch some high level attribute descriptions, which then need to get mathematically modeled. One could actually say that each of such meta-attributes corresponds to a granule of attributes which are derived using different parameters from the same model.

As a case study, we discuss the problem of MRI analysis [6], which has been already investigated in the rough set literature from the perspectives of both image segmentation [7] and tissue classification [8]. We outline several examples of meta-attributes corresponding to different ways of looking at the MRI slices. We show how to run the process of attribute selection given a huge space of attribute parameterizations, and how to avoid multiple realizations of the same meta-attributes within single reducts.

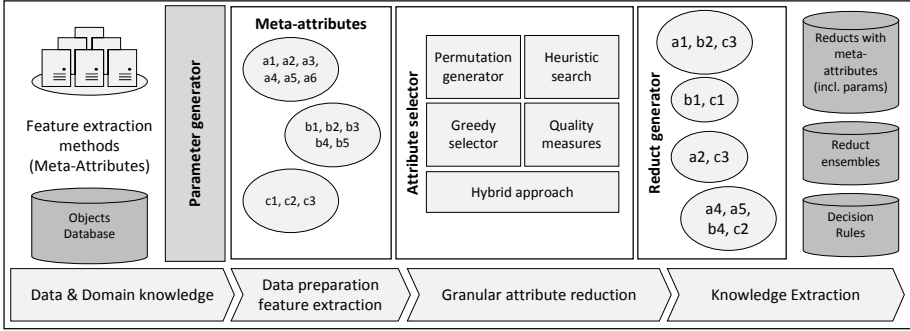


Fig. 1. Granular attribute selection framework

2 Rough Set Reduct Construction

Let us use standard notation to represent data [2]. By a decision system we mean a tuple $\mathbb{A} = (U, A \cup \{d\})$, where U is a set of objects, A is a set of attributes and $d \notin A$ is a distinguished decision attribute. We treat attributes $a \in A$ as functions $a : U \rightarrow V_a$. We say that objects $u, u' \in U$ are discerned by $a \in A$, iff $a(u)$ is different from $a(u')$. We say that $B \subseteq A$ is a decision superreduct, denoted by $B \in \mathcal{S}$, $\mathcal{S} \subseteq \mathcal{P}(A)$, iff each pair of objects discerned by d is also discerned by an element of B . We say that $B \in \mathcal{S}$ is a decision reduct, iff there are no proper subsets of B belonging to \mathcal{S} .

In practice, the criteria for being a decision superreduct can be formulated in many other ways adjusted to the types of attributes, expected level of noise in data, and so on [8]. Generally, we should work with families of approximate decision superreducts $\mathcal{S}^* \subseteq \mathcal{P}(A)$ satisfying the following two conditions: 1) $\mathcal{S} \subseteq \mathcal{S}^*$; 2) if $B \in \mathcal{S}^*$ and $B \subseteq C$, then $C \in \mathcal{S}^*$. The first condition means that standard decision superreducts should be always a special case of any of approximate superreduct extensions. The second condition means that the considered criteria are monotonic [9].

Superreduct criteria can be related to degrees of information that subsets of attributes bring about the decision. Let us consider function $I_d : \mathcal{P}(A) \rightarrow \mathcal{I}$, where (\mathcal{I}, \preceq) is a partially ordered space of information degrees. We assume that if $B \subseteq C$, then $I_d(B) \preceq I_d(C)$. We can think about approximate superreducts corresponding to certain levels of information, i.e., families $\mathcal{S}^* \subseteq \mathcal{P}(A)$ such that if $B \in \mathcal{S}^*$ and $I_d(B) \preceq I_d(C)$, then $C \in \mathcal{S}^*$. Surely, for large data sets it is important to work with functions I_d that are easy to compute. It is also useful to develop procedures for computing values of $I_d(B)$ for many subsets $B \subseteq A$ in the same time [10].

For a given function $I_d : \mathcal{P}(A) \rightarrow \mathcal{I}$ and a corresponding family $\mathcal{S}^* \subseteq \mathcal{P}(A)$, the idea of defining decision reducts is the same as before. We say that $B \in \mathcal{S}^*$ is an approximate decision reduct, iff there are no proper subsets of B belonging to \mathcal{S}^* . Fast derivation of approximate decision reducts containing possibly small amounts of attributes is highly important for representing and utilizing data dependencies. However, for decision systems with huge number of attributes, approximate decision reducts are very hard to compute. Also, it is then difficult to interpret which approximate dependencies between attributes and decisions may be meaningful for users.

Algorithm 1. Superreduct generator.

Input: meta-attributes $\{C_1, \dots, C_k\}$
Output: decision superreduct $B \in \mathcal{S}^*$

```

1  $B = \emptyset; i = 1; fail = FALSE;$ 
2 while  $B \notin \mathcal{S}^*$  or  $fail$  do
3    $fail = TRUE;$ 
4   select  $m > 0$  random parameter settings for  $C_i;$ 
5    $I_{max} = I_d(\emptyset);$ 
6    $a_{max} = NULL;$ 
7   for  $j = 1, m$  do
8     generate attribute  $c_{i_j} \in C_i$  using the  $j$ -th parameter setting;
9     if  $I_d(B \cup \{c_{i_j}\}) \neq I_{max}$  then
10       $I_{max} = I_d(B \cup \{c_{i_j}\});$ 
11       $a_{max} = c_{i_j};$ 
12       $fail = FALSE;$ 
13    end
14  end
15  if  $a_{max}$  is not  $NULL$  then
16     $B = B \cup \{a_{max}\};$ 
17     $a_{max} = NULL;$ 
18  end
19   $i = (1 - i/k) * i + 1;$ 
20 end
21 return  $B;$ 

```

Let us consider a scenario where attributes correspond to the outputs of parameterized feature extraction methods applied for some complex measurements. Such methods may generate different attributes depending on selected parameters. On the other hand, such attributes have analogous meaning for users. Reducts containing too many of such attributes might be considered as ambiguous. Figure 1 presents a framework for grouping conceptually analogous attributes into clusters and ensuring that calculated reducts do not contain multiple elements of the same clusters too often. Users can then focus on general data dependencies more easily. We call such clusters meta-attributes. Each cluster refers to a feature extraction method and its parameter generator. Thus, each meta-attribute can be represented by one or many “real” attributes.

Algorithm 1 shows how to construct an ordering of attributes corresponding to the input meta-attributes. It is a random-greedy approach based on our earlier research [5]. For each meta-attribute, we select several parameter settings and calculate values of the corresponding attributes. An attribute bringing the highest information gain in terms of a given information function $I_d : \mathcal{P}(A) \rightarrow \mathcal{I}$ is selected. We do not assume creation of a complete decision system. We calculate values of only those attributes which are selected by the algorithm. The resulting $B \in \mathcal{S}^*$, together with an ordering of added attributes, can be passed to a typical permutation based algorithm for reduct derivation [8]. However, sometimes $B \in \mathcal{S}^*$ cannot be obtained. In such cases, one should revisit expectations with respect to a reachable level of decision information.

Table 1. Examples of MRI meta-attributes: *SOM* (self-organizing map), *HC* (histogram clustering), *EDG* (edge detection filter), *NBR* (neighborhood filter), *MSK* (image mask filter)

	Description	Parameters
<i>SOM</i>	Performs unsupervised image segmentation	Number of modalities; Number of clusters; Learning rate; Number of iterations; Radius
<i>HC</i>	Performs unsupervised image segmentation based on the image histogram	Number of clusters; Bucket size; Minimum cluster distance; Peak separability; Approximation Degree; Number of clusters
<i>EDG</i>	Detects boundaries between tissues	Noise Threshold
<i>NBR_{SOM}</i>	For boundary regions assigns major <i>SOM</i> tissue class label.	Window shape; Window size
<i>NBR_{HC}</i>	For boundary regions assigns major <i>HC</i> tissue class label	Window shape; Window size
<i>MSK</i>	Encodes voxel position relative to brain central point	Number of mask regions; Region radius size; Background threshold

3 Rough Set Inspired MRI Segmentation

As a case study, let us present a rough set based approach to human brain MRI segmentation [6]. Segmentation is the process of assigning class labels to data containing spatially varying information [7]. As an example, we can consider the Simulated Brain Database (SBD) containing 3-D volumetric multi-spectral MRI images in three modalities (T1, T2, PD).¹ In this case, the goal is to classify each voxel within a brain image slice to one of tissue classes such as: cerebrospinal fluid (CSF), gray matter (GM) and white matter (WM), background (BCG), as well as others (skin, bone, fat).

In order to define meta-attributes, we can use several feature extraction methods. Combined with specific parameters, each of those methods generates an attribute labeling voxels with specific values. At the conceptual level, each meta-attribute provides some domain related information about the analyzed images. On the other hand, feature extraction parameters are often very technical, not necessarily understandable by medical experts. We need to separate those two realities and provide tools allowing users to analyze attribute dependencies without worrying about any technical settings.

The primary source of information is the voxel magnitude value registered in three MRI modalities. Using this source, we can construct more advanced attributes. In our research, we employed two unsupervised clustering algorithms: a self-organizing map (*SOM*) and our own method for image histogram discretization [8]. We also utilized the open source Insight Segmentation and Registration Toolkit (ITK)² providing a library of image filters. One can employ such filters for edge detection, voxel neighborhood analysis and relative voxel positioning against a brain center. Each of considered feature extraction methods can be applied separately to all image modalities and can be considered as a meta-attribute describing a single modality. Table 1 summarizes meta-attributes and possible parameters used within our segmentation framework.

¹ www.bic.mni.mcgill.ca/brainweb

² www.itk.org

4 Conclusions

We presented a new rough set inspired approach to attribute selection, for decision systems created using parameterized attribute value generation procedures. We described two levels of formation of attributes: meta-attributes serving as conceptual descriptions of general feature extraction methods, as well as their specific realizations constitute an actual decision system. As a case study, we presented an extension of our previous MRI image segmentation framework. In future, we will investigate usefulness of our approach also in other areas, such as predictive analytics based on sensor measurements and knowledge discovery based on text reports.

Acknowledgments. This research was supported by grant O ROB/0010/03/001 (“Modern Engineering Tools for Decision Support for Commanders of the State Fire Service of Poland during Fire & Rescue Operations in the Buildings”) founded by Polish National Centre for Research and Development, as well as grants 2011/01/B/ST6/03867 and 2012/05/B/ST6/03215 founded by Polish National Science Centre.

References

1. Kuncheva, L.I., Diez, J.J.R., Plumpton, C.O., Linden, D.E.J., Johnston, S.J.: Random Subspace Ensembles for fMRI Classification. *IEEE Transactions on Medical Imaging* 29(2), 531–542 (2010)
2. Pawlak, Z.: *Rough Sets – Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers (1991)
3. Kruczyk, M., Baltzer, N., Mieczkowski, J., Dramiński, M., Koronacki, J., Komorowski, J.: Random Reducts: A Monte Carlo Rough Set-based Method for Feature Selection in Large Datasets. *Fundamenta Informaticae* 127(1-4), 273–288 (2013)
4. Nguyen, S.H., Bazan, J., Skowron, A., Nguyen, H.S.: Layered Learning for Concept Synthesis. In: Peters, J.F., Skowron, A., Grzymała-Busse, J.W., Kostek, B.z., Swiniarski, R.W., Szczuka, M.S. (eds.) *Transactions on Rough Sets I*. LNCS, vol. 3100, pp. 187–208. Springer, Heidelberg (2004)
5. Janusz, A., Ślęzak, D.: *Rough Set Methods for Attribute Clustering and Selection*. Applied Artificial Intelligence (2014)
6. Koo, J.J., Evans, A.C., Gross, W.J.: 3-D Brain MRI Tissue Classification on FPGAs. *IEEE Transactions on Image Processing* 18(12), 2735–2746 (2009)
7. Maji, P., Pal, S.K.: Maximum Class Separability for Rough-Fuzzy C-Means Based Brain MR Image Segmentation. In: Peters, J.F., Skowron, A., Rybiński, H. (eds.) *Transactions on Rough Sets IX*. LNCS, vol. 5390, pp. 114–134. Springer, Heidelberg (2008)
8. Widz, S., Ślęzak, D.: Approximation Degrees in Decision Reduct-based MRI Segmentation. In: Howard, D., Rhee, P.K. (eds.) *Frontiers in the Convergence of Bioscience and Information Technologies 2007, FBIT 2007*, Jeju Island, Korea, October 11-13, pp. 431–436. IEEE Computer Society (2007)
9. Yao, Y., Zhao, Y., Wang, J.: On Reduct Construction Algorithms. In: Gavrilova, M.L., Tan, C.J.K., Wang, Y., Yao, Y., Wang, G. (eds.) *Transactions on Computational Science II*. LNCS, vol. 5150, pp. 100–117. Springer, Heidelberg (2008)
10. Kowalski, M., Stawicki, S.: SQL-Based Heuristics for Selected KDD Tasks over Large Data Sets. In: Ganzha, M., Maciaszek, L.A., Paprzycki, M. (eds.) *Federated Conference on Computer Science and Information Systems, FedCSIS 2012*, September 9-12, pp. 303–310. IEEE, Wrocław (2012)