Rules from granules vs. granulated rules

Krzysztof Ropiak

Faculty of Mathematics and Computer Science
University of Warmia and Mazury in Olsztyn
Poland
email:kropiak@matman.uwm.edu.pl

Abstract. This article presents a comparison of classification effects using an exhaustive set of decision-making rules and a granular set of rules. Standard approach is that we perform granulation of the chosen data set looking for the optimal granulation radius and at the end we generate new decision rules, where on the other side, our method is based on the idea of building decision rules first and then granulating them using known methods.

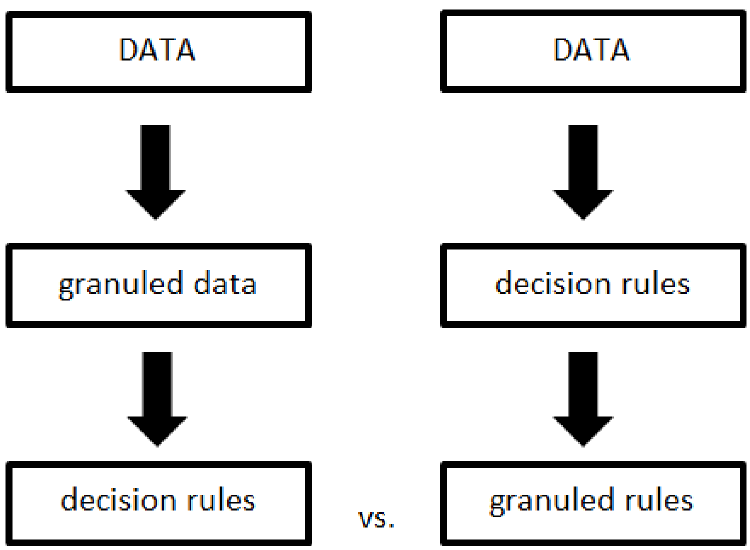

Fig. 1. Diagram visualizing comparison of both methodologies.

Keywords: data granulation, decision rules, Rough Sets, Decision Systems, Classification

1 Introduction

Data approximation methods are especially important in big data analysis where often the internal knowledge is more important than the single data sample itself. One of the most important paradigms is granular rough computing. This idea introduces the concept of data granules in terms of rough sets theory [4]. The term granule was initially used by Lotfi Zadeh [27] to define the group of objects grouped together in the sense of similarity relation.

In [10] and [11] Polkowski introduced a simple yet effective idea of data approximation using rough inclusions. This approach, named standard granulation, relies on creating granules of r-indiscernible objects, then covering of the original training data is performed and finally new objects are created from granular reflections through the use of majority voting.

New techniques and their applications were developed and described in ([1]-[3], Polkowski [9]–[14], and Polkowski and Artiemjew [17]– [24]. For other applications of data approximation process, classification and missing values absorption - see [16].

There are many other granulation techniques like concept-dependent granulation, layered granulation and recently developed homogeneous granulation [28]. Main goal of granulation process is to reduce the amount of data but at the same time the internal knowledge must be maintained to allow acceptable level of classification accuracy.

One of the simplest idea of data classification is by building decision rules based on those data and then combine them with a classifier. Once again there are many know methods of rules generation algorithms starting from exhaustive rules, by LEM2 algorithm or sequential covering . When we consider the fact, that the granulation is reducing the amount of data that is being processed during classification, and in some cases we can further reduce this amount by generating decision rules, we thought of an idea of rules granulation as an extension of this approach. In the next part of this section we will present some theoretical introduction to the classical methods we were using as a base of comparison to our approach.

The rest of the paper has the following content. In Sect. 1 we present the theoretical introduction to granular rough computing and decision rules. In Sect. 2 we detail the description of our approach to rules granulation with a toy example. In Sect. 3 we introduce the classifier used in experimental part. In Sect. 4 experiment results are presented, and we conclude the paper in Sect. 5.

The granulation process consists of three basic steps, the granules are formed around the training objects, the covering of universe of training objects is chosen, and finally granular reflection from covering granules is obtained by majority voting procedure. As a final step decision rules are built and a classification process is being performed. We begin with the basic notions of rough inclusions to introduce the first step.

1.1 Theoretical background - granular rough inclusions

The models for rough mereology which give us methods by which the rough inclusions are defined, are presented in Polkowski [6]–[10]; a detailed discussion may be found in Polkowski [15].

For a rough inclusion μ on the universe U of a decision system $D = (U, A, d)$. We introduce the parameter r_{gran}, the *granulation radius* with values $0, \frac{1}{|A|}, \frac{2}{|A|}, ..., 1$. For each object $u \in U$, and $r = r_{gran}$, the *standard granule* $g(u, r, \mu)$, *of radius* r *about* u, is defined as

$$g(u, r, \mu) \text{ is } \{v \in U : \mu(v, u, r)\}. \tag{1}$$

The standard rough inclusion is defined as

$$\mu(v, u, r) \Leftrightarrow \frac{|Ind(u, v)|}{|A|} \geq r \tag{2}$$

where

$$IND(u, v) = \{a \in A : a(u) = a(v)\}, \tag{3}$$

It follows that this rough inclusion extends the indiscernibility relation to a degree of r.

1.2 Covering of decision system

In this step the universe of training objects should be covered by computed granules using a selected strategy. One of the most effective methods among the studied ones (see [24]) is simple random choice and thus this method is selected for our experiments. In the next section there is a description of the last step of the granulation process.

1.3 Granular reflections

Once the granular covering is selected, the idea is to represent granules by single objects. The strategy for obtaining it can be the *majority voting* MV, so for each granule $g \in COV(U, \mu, r)$, the final representation is formed as follows

$$\{MV(\{a(u) : u \in g\}) : a \in A \cup \{d\}\} \tag{4}$$

where for numerical data we treat the descriptors as indiscernible in case $\frac{|a_i(u)-a_j(u)|}{max_a - min_a} \leq \varepsilon$, i, j are the numbers of objects in granule,

The granular reflection of the decision system $D = (U, A, d)$,(where U is the universe of objects, A the set of conditional attributes and d is decision attribute), $(COV(U, \mu, r))$ is formed from granules.

$$v \in g_r^{cd}(u) \text{ if and only if } \mu(v, u, r) \text{ and } (d(u) = d(v)) \tag{5}$$

for a given rough (weak) inclusion μ.

In the next section rules granulation method is presented.

2 Used method and toy example of decision rules granulation

Approach which was used for rules granulation can be described in following steps.

***Step* 1.**
Exhaustive rules from given dataset are generated.

***Step* 2.**
Rules with length = 1 (only one descriptor) are omitted and moved to final set.

***Step* 3.**
Rules are divided into separate sets with rules of the same length. For each set the concept granulation is being performed, it means that only rules from the same decision class are compared when computing the indiscernibility. New rules are created using majority voting method and possible conflicts are resolved by random.

***Step* 4.**
Newly created rules for each length and indiscernibility radius are being put together. Support (number of occurrences) for each rule is calculated and possible conflicting rules are removed. This approach favors longer rules because of the higher support value.
Because of the fact that this granulation approach brings just slightly bigger number of rules, especially for the larger datasets it is hard to present a solid toy example. Following sample from Australian-credit dataset will show the concepts of concept dependent rules granulation.

10 randomly selected rules from exhaustive set generated on Australian-credit. Rules with length 1 were omitted.

(a2=34.08) (a6=4.0) $\Rightarrow$ $d = 0.0[1]$
$(a3 = 0.25)$ $(a8 = 1.0)$ $\Rightarrow$ $d = 1.0[1]$
$(a5 = 9.0)$ $(a13 = 80.0)$ $\Rightarrow$ $d = 1.0[1]$
$(a2 = 15.83)$ $(a3 = 0.585)$ $\Rightarrow$ $d = 1.0[1]$
$(a5 = 7.0)$ $(a6 = 4.0)$ $(a14 = 2.0)$ $\Rightarrow$ $d = 0.0[1]$
$(a3 = 1.0)$ $(a9 = 1.0)$ $(a11 = 0.0)$ $\Rightarrow$ $d = 1.0[1]$
$(a7 = 0.415)$ $(a9 = 1.0)$ $(a13 = 0.0)$ $\Rightarrow$ $d = 1.0[1]$
$(a3 = 0.875)$ $(a9 = 0.0)$ $(a11 = 1.0)$ $\Rightarrow$ $d = 1.0[1]$
$(a5 = 6.0)$ $(a7 = 3.5)$ $(a10 = 0.0)$ $(a12 = 2.0)$ $\Rightarrow$ $d = 0.0[1]$
$(a1 = 1.0)$ $(a4 = 2.0)$ $(a5 = 8.0)$ $(a13 = 160.0)$ $(a14 = 1.0)$ $\Rightarrow$ $d = 1.0[1]$

Let's take three rules into consideration
$(a3 = 1.0)$ $(a9 = 1.0)$ $(a11 = 0.0)$ $\Rightarrow$ $d = 1.0[1]$
$(a7 = 0.415)$ $(a9 = 1.0)$ $(a13 = 0.0)$ $\Rightarrow$ $d = 1.0[1]$
$(a3 = 0.875)$ $(a9 = 0.0)$ $(a11 = 1.0)$ $\Rightarrow$ $d = 1.0[1]$

These are all rules with length=3 and decision=1 so those rules are granuled as single set. As we can see two of them have the same attribute numbers with slightly different values. When we run a majority voting on those two rules a new rule will be built because the dominant value is the same for each attribute and random choice will be used. At this run a new rule
$(a3 = 1.0)$ $(a9 = 0.0)$ $(a11 = 0.0)$ $\Rightarrow$ $d = 1.0[1]$
was built, but this process is not deterministic.

Final set of granuled rules looks as follows.
$(a2 = 34.08)$ $(a6 = 4.0)$ $\Rightarrow$ $d = 0.0[3]$
$(a3 = 0.25)$ $(a8 = 1.0)$ $\Rightarrow$ $d = 1.0[3]$
$(a5 = 9.0)$ $(a13 = 80.0)$ $\Rightarrow$ $d = 1.0[3]$
$(a2 = 15.83)$ $(a3 = 0.585)$ $\Rightarrow$ $d = 1.0[3]$
$(a5 = 7.0)$ $(a6 = 4.0)(a14 = 2.0)$ $\Rightarrow$ $d = 0.0[4]$
$(a3 = 1.0)$ $(a9 = 0.0)(a11 = 0.0)$ $\Rightarrow$ $d = 1.0[1]$
$(a7 = 0.415)$ $(a9 = 1.0)(a13 = 0.0)$ $\Rightarrow$ $d = 1.0[4]$
$(a3 = 0.875)$ $(a9 = 0.0)(a11 = 1.0)$ $\Rightarrow$ $d = 1.0[3]$
$(a3 = 1.0)$ $(a9 = 1.0)(a11 = 0.0)$ $\Rightarrow$ $d = 1.0[3]$
$(a5 = 6.0)$ $(a7 = 3.5)(a10 = 0.0)(a12 = 2.0)$ $\Rightarrow$ $d = 0.0[5]$
$(a1 = 1.0)$ $(a4 = 2.0)$ $(a5 = 8.0)$ $(a13 = 160.0)$ $(a14 = 1.0)$ $\Rightarrow$ $d = 1.0[6]$

3 Classification process

We have designed rule based classifier, which consists of fitting the set of decision rules into classified objects and vote for decision. The ties are resolved randomly. We have measured the effectiveness using global accuracy parameter, which can be defined as the percentage of correctly classified objects. We use exhaustive set of rules in our experiments, with minimal descriptors length. We are generating no conflicting decision rules starting from the ones with length equal to one. This algorithm finishes his work with the rule length, in which there are no more candidates or minor number of rules, in comparison with the whole computed set, is generated.

4 Results

We have carried out experiments on exemplary real data from UCI Repository. Used dataset description is presented in Table 1 while results of granulation and classification in Table 2 and Table 3 respectively.

Table 1. Data Sets description - see [26]

name	*attr type*	*attr no.*	*obj no.*	*class no.*
$Australian - credit$	$integer,\ real$	15	690	2
$Iris$	$integer,\ real$	4	150	3
$Diabetes$	$integer,\ real$	9	768	2
$Liver$	$integer,\ real$	6	345	2

Table 2. Comparison of number of rules in both approaches

name	*# exhaustive rules*	*# granuled rules*	*abs. diff.*	*% diff.*
$Australian - credit$	12025	13804	1779	14.79%
$Iris$	246	247	1	0.41%
$Diabetes$	8609	8786	177	2.06%
$Liver$	2844	2918	74	2.60%

Table 3. Classification results - average accuracy after cv5

name	*exhaustive rules*	*granuled rules*
$Australian - credit$	84.20%	82.32%
$Iris$	85.89%	86.67%
$Liver$	57.10%	56.52%
$Diabetes$	65.37%	61.72%

We have used our own implementation of exhaustive algorithm, basic method for results evaluation is Cross Validation 5 technique.

5 Conclusions

In this work we have considered the technique of exhaustive set of rules granulation. We have compared the set of decision rules after their granulation vs rules computed from granulated decision systems. It was proven that its better to granulate and compute rules than compute rules and granulate them. In the latter case we have lose of the information. It is difficult to merge rules after their granulation. The granulation process of exhaustive decision rules seems to be ineffective, because the rules are designed in MDL (minimal description length) model, and they are not redundant. Granulation process works good in case there are many indiscernible values in the granulated entity.

6 Acknowledgements

The research has been supported by grant 23:610:007-300 from Ministry of Science and Higher Education of the Republic of Poland.

References

1. Artiemjew, P. : Classifiers from Granulated Data Sets: Concept Dependent and Layered Granulation. In Proceedings RSKD'07. The Workshops at ECML/PKDD'07, Warsaw Univ. Press, Warsaw, 2007, pp 1–9 (2007)
2. Artiemjew, P.: Natural versus granular computing: Classifiers from granular structures. In Proceedings of 6th International Conference on Rough Sets and Current Trends in Computing RSCTC'08, Akron OH, USA, (2008)
3. Artiemjew, P. (2013): A Review of the Knowledge Granulation Methods: Discrete vs. Continuous Algorithms. In Skowron A., Suraj Z. (eds.)(2013): Rough Sets and Intelligent Systems. ISRL 43, Springer-Verlag, Berlin, 2013, pp 41–59.
4. Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 11, pp 341–356 (1982)
5. Polap, D., Wozniak, M., Wei, W., Damasevicius, R.: Multi-threaded learning control mechanism for neural networks. Future Generation Computer Systems, Elsevier 2018.
6. Polkowski, L.: Rough Sets. Mathematical Foundations. Physica Verlag, Heidelberg (2002)
7. Polkowski, L.: A rough set paradigm for unifying rough set theory and fuzzy set theory. Fundamenta Informaticae 54, pp 67–88; and : In Proceedings RSFDGrC03, Chongqing, China, 2003. Lecture Notes in Artificial Intelligencevol. 2639, Springer Verlag, Berlin, pp 70–78 (2003)
8. Polkowski, L.: Toward rough set foundations. Mereological approach. In Proceedings RSCTC04, Uppsala, Sweden. Lecture Notes in Artificial Intelligence vol. 3066, Springer Verlag, Berlin, pp 8–25 (2004)
9. Polkowski, L.: Granulation of knowledge in decision systems: The approach based on rough inclusions. The method and its applications In Proceedings RSEISP'07, Lecture Notes in Artificial Intelligence vol. 4585. Springer Verlag, Berlin, pp 69– (2004)

10. Polkowski, L.: Formal granular calculi based on rough inclusions. In Proceedings of IEEE 2005 Conference on Granular Computing GrC05, Beijing, China. IEEE Press, pp 57–62 (2005)
11. Polkowski, L.: A model of granular computing with applications. In Proceedings of IEEE 2006 Conference on Granular Computing GrC06, Atlanta, USA. IEEE Press, pp 9–16 (2006)
12. Polkowski, L.: The paradigm of granular rough computing. In Proceedings ICCI'07, Lake Tahoe NV. IEEE Computer Society, Los Alamitos CA, pp 145–163 (2007)
13. Polkowski, L.: A Unified Approach to Granulation of Knowledge and Granular Computing Based on Rough Mereology: A Survey, in: Handbook of Granular Computing, Witold Pedrycz, Andrzej Skowron, Vladik Kreinovich (Eds.), John Wiley & Sons, New York, 375-401 (2008)
14. Polkowski, L.: Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems. In Meyers, R. A.(ed.): Encyclopedia of Complexity and System Sciences. Springer Verlag, Berlin, article 00788 (2009)
15. Polkowski, L.: Approximate Reasoning by Parts. An Introduction to Rough Mereology. Springer Verlag, Berlin, (2011)
16. Polkowski, L., Artiemjew, P.: On granular rough computing with missing values. In Proceedings RSEISP'07, Lecture Notes in Artificial Intelligence vol. 4585. Springer Verlag, Berlin, 2007, pp 271–279 (2007)
17. Polkowski, L., Artiemjew, P.: On granular rough computing: Factoring classifiers through granular structures. In Proceedings RSEISP 2007, Lecture Notes in Artificial Intelligence vol. 4585. Springer Verlag, Berlin, pp 280–290 (2007)
18. Polkowski, L.,Artiemjew, P.: Towards Granular Computing: Classifiers Induced from Granular Structures. In Proceedings RSKD'07. The Workshops at ECML/PKDD'07, Warsaw Univ. Press, Warsaw, pp 43–53 (2007)
19. Polkowski, L., Artiemjew, P.: Classifiers based on granular structures from rough inclusions. In Proceedings of 12th Int. Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU'08, Torremolinos (Malaga), Spain, pp 1786–1794 (2008)
20. Polkowski, L.,Artiemjew P.: Rough sets in data analysis: fundations and applications. In Smoliński, T. G., Milanova, M., Hassanien, A.–E. (eds.): Applications of Computional Intelligence in Biology: Current Trends and open Problems, SCI, vol. 122. Springer Verlag, Berlin, 2008, pp 33–54.
21. Polkowski, L., Artiemjew, P.: Rough mereology in classification of data: Voting by means of residual rough inclusions. In Proceedings of 6th International Conference on Rough Sets and Current Trends in Computing RSCTC'08, Akron OH, USA. Lecture Notes in Artificial Intelligence vol. 5306, Berlin, pp 113–120 (2008)
22. Polkowski, L., Artiemjew, P.: A study in granular computing: On classifiers induced from granular reflections of data. Transactions on Rough Sets IX. Lecture Notes in Computer Science vol. 5390. Springer, Berlin, pp 230–263 (2008)
23. Polkowski, L., Artiemjew, P.: On classifying mappings induced by granular structures. Transactions on Rough Sets IX. Lecture Notes in Computer Science vol. 5390. Springer, Berlin, pp 264–286 (2008)
24. Polkowski, L., Artiemjew, P.: Granular Computing in Decision Approximation - An Application of Rough Mereology, in: Intelligent Systems Reference Library 77, Springer, ISBN 978-3-319-12879-5, 1-422 (2015)
25. Ohno-Machado, L.: Cross-validation and Bootstrap Ensembles, Bagging, Boosting, Harvard-MIT Division of Health Sciences and Technology,

http://ocw.mit.edu/courses/health-sciences-and-technology/hst-951j-medical-decision-support-fall-2005/lecture-notes/hst951_6.pdf HST.951J: Medical Decision Support, Fall (2005)
26. University of California, Irvine Machine Learning Repository: https://archive.ics.uci.edu/ml/index.php
27. Zadeh, L. A.: Fuzzy sets and information granularity. In Gupta, M., Ragade, R., Yager, R.R. (eds.): Advances in Fuzzy Set Theory and Applications. North–Holland, Amsterdam, 1979, pp 3–18 (1979)
28. Ropiak, K., Artiemjew, P.: A Study in Granular Computing: homogenous granulation. In: Dregvaite G., Damasevicius R. (eds) Information and Software Technologies. ICIST 2018. Communications in Computer and Information Science, Springer (2018)