On Granular Rough Computing: Covering by Joint and Disjoint Granules in Epsilon Concept Dependent Granulation

Piotr Artiemjew and Jacek Szypulski

Abstract In this work we present the optimization methods of epsilon concept-dependent granulation. We consider two cases of parallel covering and granulation, based on joint and disjoint granules. Additionally we check two variants of majority voting, the first one based on descriptors, which are epsilon-indiscernible with the centers of granules, and the second variant uses all descriptors of respective granules. We verify the effectiveness of our methods on the real data sets from UCI Repository using the SVM classifier. It turned out that disjoint granules versus joint give almost identical results of classification with a significant acceleration of the granulation process. Additionally, the majority voting, based on the epsilon indiscernible descriptors, stabilised the process of granulation in terms of the accuracy of classification. This is a significant result, which lets us to accelerate the process of classification for many popular classifiers at least for k-NN, Naive Bayes, many rough set methods and the SVM classifier, which is supported by our recent works.

Keywords Rough sets · Decision systems · SVM · Granular rough computing · Epsilon concept-dependent granulation · Majority voting

1 Introduction

In recent years the granular computing approach has gained great interest among researchers. The popularity of the approach can be explained by the analogy to natural thinking; it is obvious that we group objects using some similarity measures, by some joint features. We use the granulation of knowledge to resolve problems in ordinary life.

P. Artiemjew (✉) · J. Szypulski
Department of Mathematics and Computer Science,
University of Warmia and Mazury, Sloneczna 54, 10-710 Olsztyn, Poland
e-mail: artem@matman.uwm.edu.pl

J. Szypulski
e-mail: jszypulski@matman.uwm.edu.pl

I. Czarnowski et al. (eds.), *Intelligent Decision Technologies 2016*,
Smart Innovation, Systems and Technologies 57,
DOI 10.1007/978-3-319-39627-9_13

The granular rough computing paradigm was initiated by Professor Lotfi Zadeh in 1979. The paradigm is connected with the Rough Set Theory proposed by Professor Zdzisław Pawlak in 1982. In the rough set theory the granules are defined as indiscernibility classes, where as similarity measures we use rough inclusions. In terms of rough inclusions, an interesting way of granulation of knowledge was proposed by Polkowski in [5, 6]. These methods turned out to be effective in many contexts. Interesting results can be found in the [7, 8]. In this scope of methods, the standard granulation is the basic one—see [5, 6]—this method was extended in joint works of Polkowski and Artiemjew into other variants, among others into granulation in the range of decision classes, and with epsilon variant, considering the indiscernibility ratio of descriptors epsilon (ε)—see [1, 2].

In this paper we have examined a few methods of optimization of the mentioned epsilon concept-dependent granulation. We propose the methods which let us compute the granules and covering of the universe of objects in the parallel way. In the covering process, we use two types of granules, the joint and disjoint one—see the Sects. 2.1 and 2.2. In the process of granular reflection creation, we use two variants of majority voting, with either use of all descriptors in the granules, or only the epsilon indiscernible to centers of granules—see Sect. 2.3. In order to compare mentioned methods we have designed an experimental session on the data from UCI Repository and with the use of the SVM classifier [4].

The rest of the paper is as follows. In Sects. 1.1, 1.2, and 1.3, we have described the methodology, the theoretical introduction to granulation in rough mereology, and the basic information about the used classifier, respectively. In Sect. 2. we have a description of our modifications of epsilon concept-dependent granulation. In Sect. 3 we have an experimental session with the results. In Sect. 4 we have conclusions and future work.

Let us start with a brief description of the methodology.

1.1 Methodology

We use the SVM classifier with RBF kernel as a reference classifier. The motivation to use it arises from our recent experiments [9], which show the effectiveness of this classifier in the context of epsilon concept-dependent granulation. Our modification of the granulation consists of the modification of covering process, which is parallel with the granulation. We have two variants. In the first, the granules are created from redundant indiscernibility classes, and in the second from disjoint indiscernibility classes. In both methods we consider only the central objects, which are new for the covering set. For verification of results we compute the accuracy of classification with the use of the five times Cross Validation 5 method [10].

In the next subsection we show the background information of our methods.

1.2 Granulation in Rough Mereology

Rough mereology is a theory of the predicate $\mu(x, y, r)$ read: "x is a part of y to a degree r", called a *rough inclusion*, see [7].

We recall that an *information system* (a *data table*) is represented as a pair (U, A) where U is a finite set of things and A is a finite set of *attributes*; each attribute $a : U \to V$ maps the set U into the *value set* V. For an attribute a and a thing v, $a(v)$ is the value of a on v.

We apply a particular form of a rough inclusion defined as follows.

For an attribute a, we let a_{max}, a_{min}, the maximal, resp. the minimal value of the attribute a on objects in the decision system, and then $span(a) = a_{max} - a_{min}$ is the span of a.

Given a parameter ε, defined as the fraction $x \times span(a)$ for $x = 0, 1, 2, \ldots, 100$ percent, we fix a *granulation radius* r.

We call two values $a(u), a(v)$ of the attribute a *ε–similar* if the inequality $\frac{|a(u)-a(v)|}{span(a)} \leq \varepsilon$ holds, in symbol $sim_\varepsilon(a(u), a(v))$. For a given object u, we define the *granule about u and of the radius r*, $g_\varepsilon(u, r)$ as the set,

$$g_\varepsilon(u, r) = \{v : |\{a \in A : sim_\varepsilon(a(u), a(v))\}| \geq r\}, \tag{1}$$

where $|.|$ denotes the size of a set.

Having granules defined, we continue with the granulation procedure. We apply the sequential covering method by selecting an object, building a granule around it, removing the granule from the universe of objects, and repeating until all objects are covered. Each of the obtained granules in the covering is factorized by selecting for each attribute the representative value for the granule by majority voting with random tie resolution. By this process, each granule is replaced with a vector of attribute values. The obtained reflection of the original decision system is then subject to classification by means of C–SVC, with the radial basis kernel RBF, see [3, 4].

1.3 The Classifier in a Nutshell

As a reference classifier we use the Support Vector Machine classifier with RBF kernel [3, 4], which turns out to be effective in the context of classification of the granular reflections of data [9]. The training and test data are normalized into the interval $[-1, 1]$, and then the classifier is used on granulated training parts of data sets.

The granulation methods used in the work are as follows.

2 Optimisation of Epsilon Concept Dependent Granulation

2.1 Epsilon Concept Dependent Granulation with Disjoint Granules

This method is the modification of [1, 2]. In this variant we fix the epsilon parameter—the descriptors discernibility ratio—and we compute the indiscernibility classes from the universe of objects. We choose only classes whose central objects are not yet in the covering, and the granules cannot contain any objects from the covering.

The detailed procedure for covering is the following,

(i) from the original decision system (U, A, d), we form the training decision system (TRN) and test decision system (TST),
(ii) $U_{cover} = \emptyset$,
iii we set the granulation radius r_{gran} and the indiscernibility ratio of attributes ε,
(iv) for given $TRN = \{u_1, u_2, ..., u_{|TRN|}\}$, we form the $TRN_{temp} = TRN - U_{cover}$, we get in a random way the object $u \in TRN_{temp}$, and form the granule

$$g_{r_{gran}}^{\varepsilon,cd}(u) = \{v \in TRN_{temp}; \frac{|IND_{\varepsilon}(u,v)|}{|A|} \geq r_{gran} \; and \; d(u) = d(v)\}$$

$$IND_{\varepsilon}(u,v) = \{a \in A : \frac{|a(u) - a(v)|}{span(a)} \leq \varepsilon\}$$

(v) $U_{cover} \leftarrow g_{r_{gran}}^{\varepsilon,cd}(u)$,
(vi) if the U_{cover} is equal TRN, we go to (vii), otherwise to (iv),
(vii) we form the granular reflections of the original TRN system based on the granules from U_{cover} with the use of selected majority voting strategy—see Sect. 2.3.

2.2 Epsilon Concept Dependent Granulation with Joint Granules

This is the modification of the previous method, where during the covering process we use indiscernibility classes computed in the entire TRN set. The central objects of new granules are still on the outside of U_{cover}. The procedure is analogous to the previous one with the exception of the granule definition, which is as follows.

$$g_{r_{gran}}^{\varepsilon,cd}(u) = \{v \in TRN; \frac{|IND_{\varepsilon}(u,v)|}{|A|} \geq r_{gran} \; and \; d(u) = d(v)\}$$

2.3 Majority Voting with Consideration of ε-indiscernible Descriptors

In the paper we consider two variants of granular reflections. We form the granular reflections from granules of covering, based on the descriptors, which are ϵ-indiscernible from the central objects of granules. The covering of the universe U is as follows.

$Cov(U) = \{g^{\epsilon,cd}_{r_{gran}}(x_i) : \bigcup_{i=1}^{k} g^{\epsilon,cd}_{r_{gran}}(x_i) = U\}$.

If $|g^{\epsilon,cd}_{r_{gran}}(x_i)| = n$ then,

$$g^{\epsilon,cd}_{r_{gran}}(x_i) = \begin{pmatrix} a_1(x_1) & a_2(x_1) & ... & a_m(x_1) \\ a_1(x_2) & a_2(x_2) & ... & a_m(x_2) \\ ... & ... & ... & ... \\ a_1(x_n) & a_2(x_n) & ... & a_m(x_n) \end{pmatrix}$$

Considering the central object of the granule,

$$a_1(x)a_2(x)...a_m(x)$$

we have to perform the following procedure of MV_{type2}.

$$MV_{type2}(a_1(g^{\epsilon,cd}_{r_{gran}}(x_i))) = Avg\{a_1(x_j) : ||a_1(x) - a_1(x_j)|| \leq \varepsilon \text{ and } x_j \in g^{\epsilon,cd}_{r_{gran}}(x_i)\}$$

$$||a_1(x) - a_1(x_j)|| \leq \varepsilon \text{ if } \frac{|a_1(x) - a_1(x_j)|}{span(a_1)} \leq \varepsilon$$

And the granular reflection of $g^{\epsilon,cd}_{r_{gran}}(x_i)$ looks as follows:

$$MV_{type2}(a_1(g^{\epsilon,cd}_{r_{gran}}(x_i))), MV_{type2}(a_2(g^{\epsilon,cd}_{r_{gran}}(x_i))), \ldots, MV_{type2}(a_m(g^{\epsilon,cd}_{r_{gran}}(x_i)))$$

The MV_{type1} is just the averaging of all of the descriptors in the granule from the covering.

$$MV_{type1}(a_1(g^{\epsilon,cd}_{r_{gran}}(x_i))) = Avg\{a_1(x_j); x_j \in g^{\epsilon,cd}_{r_{gran}}(x_i)\}$$

3 Experimental Session

For the experimental session we use the covering with hierarchical choice and the same split of data sets for a better comparison of methods. As a reference classifier we use SVM with the multiple Cross Validation method. The training system of each split was granulated by the respective method. In the experiments we

use selected data from UCI Repository, among others, *Wisconsin Diagnostic Breast Cancer*, *Breast Cancer Wisconsin*, *Wisconsin Prognostic Breast Cancer*, *Fertility*, and *Parkinsons*.

On the charts, we have the Average accuracy of classification—this is the average accuracy from 5 times CV-5 test. There is the parameter Radius (r_{gran}) the granulation radius and Epsilon (ε) the indiscernibility ratio of descriptors. These parameters are useful in the process of approximation. The average size of the training set (granular decision system) is the percentage size of the training decision system after the approximation process. These parameters, compared with the accuracy of classification, show us the level of acceleration of the classification process.

3.1 Results of Experiments

Due to lack of space we show the result only for the exemplary data set, but the conclusion of our work is consistent with all of our results.

The results for the chosen *Wisconsin Diagnostic Breast Cancer* data set are as follows. In Fig. 1, we have the accuracy of classification for joint granules and majority voting based on the epsilon indiscernible descriptors. In Fig. 2 we have the analogous result for majority voting with the use of all descriptors of granules. Furthermore, in Figs. 3, and 4, we have the respective results for the disjoint granules. Finally, in Fig. 5 we have the average size of the granulated data set, which lets us see the possible acceleration of classification. We can see the spectra of parameters, which allows us to preserve knowledge from the original training data set.

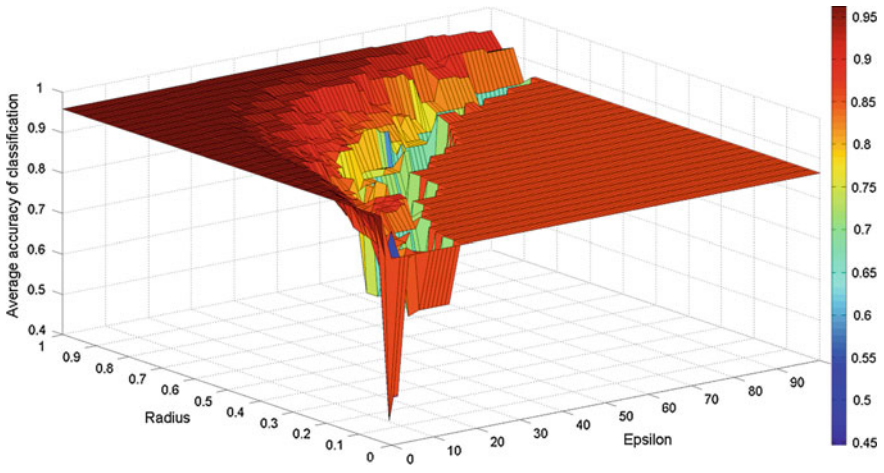

Fig. 1 5xCV5—classification result for WDBC data set for epsilon concept-dependent granulation and granules created from redundant indiscernibility classes; the averaging of values in majority voting with considering descriptors, which are epsilon indiscernible with central objects of granules

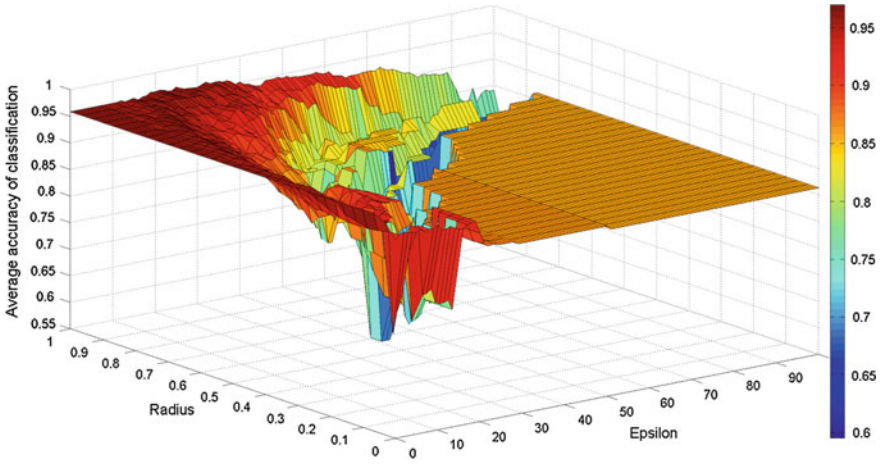

Fig. 2 5xCV5—classification result for WDBC data set for epsilon concept-dependent granulation and granules created from redundant indiscernibility classes; the averaging of values in majority voting with considering all descriptors

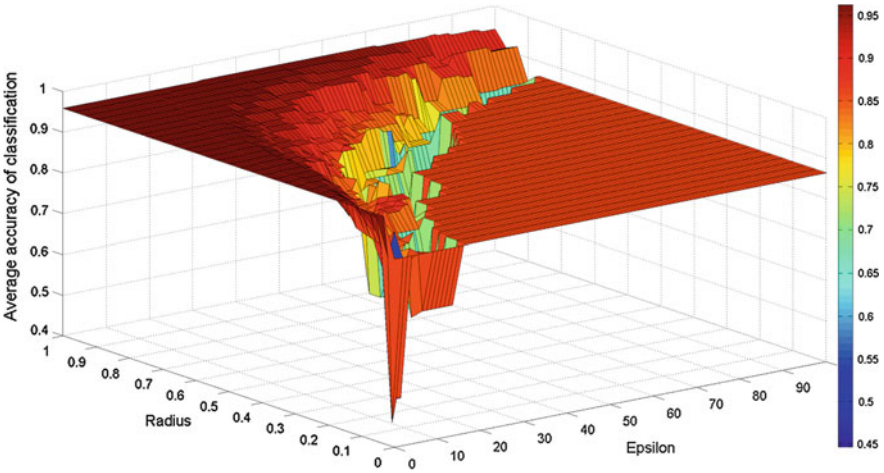

Fig. 3 5xCV5—classification result for WDBC data set for epsilon concept-dependent granulation and granules created from disjoint indiscernibility classes; the averaging of values in majority voting with considering descriptors, which are epsilon indiscernible with central objects of granules

The results of the experiments show the advantage of the method based on the disjoint granules. For both cases of majority voting, the result is comparable with the result for joint granules, but in the first method we have significant acceleration of granulation. Comparing the majority voting strategies, the one based on the epsilon indiscernible descriptors stabilises the classification process.

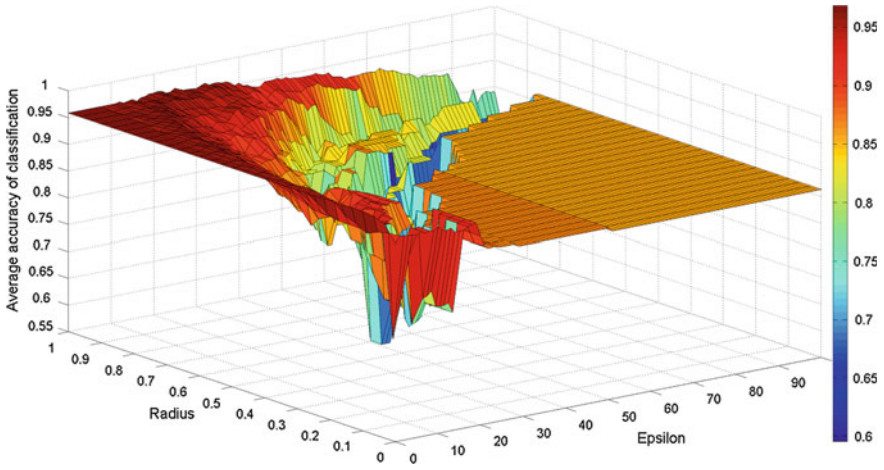

Fig. 4 5xCV5—classification result for WDBC data set for epsilon concept-dependent granulation and granules created from disjoint indiscernibility classes; the averaging of values in majority voting with considering all descriptors

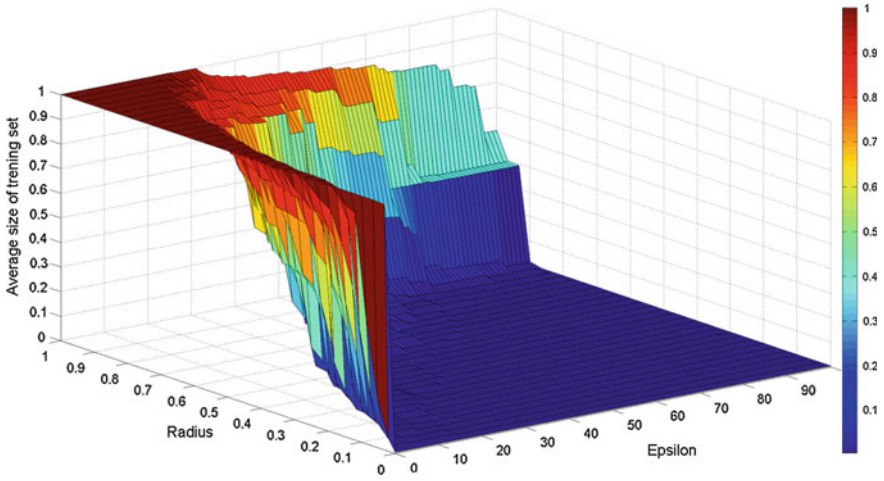

Fig. 5 The average size of granular decision systems for concept-dependent granulation variants with WDBC data set

The best results here are the ones with good enough accuracy and high lowering of the training data set size. In those cases the approximation of the training data set yield significant maintenance of knowledge from the original training data set. This was proven in many previous works, for example: [8, 9].

On the question 'why does the method of granulation based on the disjoint granules work in most cases faster than the one based on the joint granules?' one can answer that indiscernibility classes overlap in many cases (for many parameters), at

least for the best parameters, which leads to high approximation of TRN data with quite high effectiveness of classification. Using joint granules we use such overlapping indiscernibility classes one by one during the covering of the universe, and we have to perform redundant operations. Thus using disjoint granules, we use indiscernibility classes with disjoint centers during covering. The process of covering speeds up because convergence occurs more quickly. This is the main source of acceleration, but one can see that the level of acceleration depends on the internal logic of the data sets, and could depend on the density of the data.

4 Conclusions

The basic result of this work is the acceleration of the granulation process by computing disjoint granules in comparison with the joint variant. The result of classification for these methods is almost identical.

Additionally, we have investigated two methods of majority voting for granular reflections creation. The results of the experiments lead us to the conclusion that the majority voting with consideration of only the ε indiscernible descriptors stabilises the granulation in the sense of accuracy of classification. The disadvantage of this method is the need for selection of the mentioned indiscernible descriptors during the granulation process, but the majority voting procedure is accelerated.

An accelerated process of granulation gives us the acceleration of classification for any classifier based on the approximated data set. In particular, in this work we can see the acceleration of the SVM classifier.

In future work we would like to check the other variants of majority voting, especially the voting on decision based on the weights determined by ε indiscernible descriptors.

Acknowledgments The research has been supported by grant 1309-802 from the Ministry of Science and Higher Education of the Republic of Poland.

References

1. Artiemjew, P.: On strategies of knowledge granulation and applications to decision systems. Polkowski, L., Supervisor, Ph.D. Dissertation, Polish Japanese Institute of Information Technology, Warsaw (2009)
2. Artiemjew, P.: A review of the knowledge granulation methods: discrete versus continuous algorithms. In: Rough Sets and Intelligent Systems—Professor Zdzisław Pawlak in Memoriam, vol. 2, pp. 41–59. Springer, Heidelberg (2013)
3. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92, pp. 144–152. ACM, New York, NY, USA (1992)
4. Chang, C.C., Lin, C.J.: Libsvm: A library for support vector machines. ACM Trans. Intell. Syst. Technol. **2**(3), 27:1–27:27 (2011)

5. Polkowski, L.: Formal granular calculi based on rough inclusions. In: 2005 IEEE International Conference on Granular Computing, vol. 1, pp. 57–69 (2005)
6. Polkowski, L.: Granulation of knowledge in decision systems: the approach based on rough inclusions. The method and its applications. In: Proceedings of International Conference on Rough Sets and Intelligent Systems Paradigms, RSEISP 2007, Warsaw, Poland, 28–30 June 2007, pp. 69–79. Springer, Heidelberg (2007)
7. Polkowski, L.: Approximate Reasoning by Parts: An Introduction to Rough Mereology, 1st edn. Springer Publishing Company (2011)
8. Polkowski, L., Artiemjew, P.: Granular Computing in Decision Approximation: An Application of Rough Mereology. Springer Publishing Company (2015)
9. Szypulski, J., Artiemjew, P.: The rough granular approach to classifier synthesis by means of SVM. In: Proceedings of the 15th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, RSFDGrC 2015. Tianjin, China, 20–23 Nov 2015, pp. 256–263. Springer International Publishing, Cham (2015)
10. Tibshirani, R.J., Tibshirani, R.: A bias correction for the minimum error rate in cross-validation. Ann. Appl. Stat. **3**(2), 822–829 (2009)