# Classification Approach based on Rough Mereology

Mahmood A. Mahmood, Nashwa El-Bendary, Aboul Ella Hassanien, and Hesham A. Hefny

**Abstract** This article presents a classification approach based on granular computing combined with rough set. The proposed classification approach used the theory of rough mereology and fuzzification in order to classify input datasets into sets of optimized granules. The proposed approach was applied to five datasets of the UC Irvine Machine Learning Repository. The Abalone dataset that consists of 4177 objects and eight attributes was selected as an illustrative example. Empirically obtained experimental results demonstrated that the proposed rough mereology based classification approach obtained better performance compared to other experienced proposed classification approaches.

Mahmood A. Mahmood
ISSR, Computer Sciences and Information Dept., Cairo University, Cairo - Egypt,
Scientific Research Group in Egypt (SRGE), http://www.egyptscience.net
e-mail: mahmood.moneim@egyptscience.net

Nashwa El-Bendary
Arab Academy for Science,Technology, and Maritime Transport, Cairo - Egypt,
Scientific Research Group in Egypt (SRGE), http://www.egyptscience.net
e-mail: nashwa.elbendary@ieee.org

Aboul Ella Hassanien
Information Technology Dept., Faculty of Computers and Information, Cairo University, Cairo - Egypt,
Scientific Research Group in Egypt (SRGE), http://www.egyptscience.net
e-mail: aboitcairo@gmail.com

Hesham A. Hefny
ISSR, Computer Sciences and Information Dept., Cairo University, Cairo - Egypt
e-mail: h.hefny@ieee.org

## 1 Introduction

Granular computing is a methodology that emulates human brain for solving different problems in the real world. Humans usually solve problems by dividing them into sub-problems then solve each sub-problem separately in order to reach the solution of the main problem as the result of all sub-problems solutions integrated together. Granular computing represents a hierarchical structure, like tree or graph, so that the root of this structure is the main problem and each level consists of some granules and the last level contains the leaf granules. One main classifier that is defined under granular computing is the theory of rough mereology, which is a paradigm for reasoning under uncertainty with primitive notion as a part to a degree. Complexity time of indiscernibility of rough mereology is non-linear, so rough inclusion technique is an efficient way to compute the indiscernibility of rough mereology via linear complex time.

Some recent research work that applied granular computing, rule generation, and rough mereology approaches are done by T.Y.Lin et al. in [1] and Hong Zhen Zheng et al. in [2] presented bitmap based association rule algorithms using granular computing technique. Authors in both researches indicated that bitmap and granular computing techniques can greatly enhance the performance for finding association rule. Moreover, Denise and Jean in [3] proposed an extension for the conventional upper approximation, termed $upper^{\alpha}$ approximation, that restricts the elements of the region of transition between a given class and its complement with respect to an $\alpha - cut$. Authors used that new concept for utilizing fuzzy rules to compose a fuzzy classifier. Furthermore, Jiujiang et al. in [4] presented a rule generation algorithm based on granular computing (RGAGC) that generates rule from the granule space without considering the selection of attributes like many classic decision tree methods. Hence, RGAGC approach used uncertainty support method to build granules, while the previously presented approaches used the ID3 (Iterative Dichotomiser 3) algorithm to divide the information into granules.

This article presents a classification approach based on granular computing combined with rough set. The proposed classification approach used the theory of rough mereology and fuzzification in order to classify input datasets into sets of generated rules in three phases; namely pre-processing, clustering, and voting by objects. For pre-processing phase, the input dataset was mapped into a normalized dataset of the range [0,1], where the values 0 and 1 represent the smallest the largest values in each dataset's attribute, respectively. For clustering phase, theory of rough mereology and rough inclusion algorithm were used for classifying the resulted dataset into sets of granules with different radius. Finally, for voting by objects phase, voting by objects algorithm was applied to select the optimized granules of dataset. The proposed approach was applied to five datasets of the UC Irvine Machine Learning Repository. The Abalone dataset that consists of 4177 objects and eight attributes was selected as an illustrative example. The rest of this article is organized as follows. Section 2 presents an overview about rough mereology approach. Section 3 describes the proposed rough mereology based classification approach. Section 4 introduces and discusses experimental results. Finally, section 5 addresses conclusions and discusses future work.

## 2 Rough Mereology: An Overview

Rough mereology can be classified according to the measurement of similarity. The similarity functions [5] that satisfy certain similarity properties, namely Monotonic (MON), Identity (ID), Extreme, or proportionality (EXT), can be defined as follows:

- (MON) if similarity (x, y, 1) then for each z, where $d(z,x) > d(z,y)$, from similarity (z, x, r) it follows that similarity (z, y, r).
- (ID) similarity (x, x, 1) for each x.
- (EXT) if similarity (x, y, r) and $s \leq r$ then similarity (x, y, s).

Rough mereology proposed by Lesniewski in [6] as the theory of concept, where the relation of mereology is a part of relation, e.g. $x$ mereology $y$ means $x$ is a part of $y$, according to Polkowski [7], the mereology relation described in equation (1)

$$ing(u,w) \Leftrightarrow \pi(u,w) \ or \ u = w \tag{1}$$

where $\pi(u,w)$ is a partial relation (proper part) and $ing(u,w)$ is ingredient relation means an improper part. $\mu(x,y,r)$ means rough mereology relation $x$ is part of $y$ at least degree $r$, also described as shown in equation (2):

$$\mu(x,y,r) = sim_{\delta}(x,y,r) \Leftrightarrow \rho(x,y) \leq (1 - -r) \tag{2}$$

Computing the indiscernibility relation to get the object can be achieved by using rough inclusion, which is of less complexity time than the indiscernibility relation computed by rough set technique. Rough inclusion from metric, according to Polkowski in [7], can be computed by the Euclidean metric space or Manhattan space, where

$$\mu_h(x,y,r) \Leftrightarrow \rho(x,y) \leq 1 - -r$$

Rough inclusion technique satisfies the similarity properties. Datasets are formalized as decision systems of the form of triple (U, A, d) or information system of the form (U, A), where $U$ is a finite set of objects, $A$ is a finite set of attributes, each attribute $a \in A$ described as mapping $a : U \rightarrow V_a$ of objects in $U$ into the value set of $a$, and $d \notin A$ is the decision. The indiscernibility relation $Ind$ can be computed as shown in equation (3):

$$Ind(x,y) = \frac{|IND(x,y)|}{|A|} \tag{3}$$

Then equation (2) becomes:

$$\mu_h(x,y,r) \Leftrightarrow Ind(x,y) \geq r \tag{4}$$

and

$$IND(x,y) = a \in A : a(x) = a(y) \tag{5}$$

where $a$ is an attribute(s) in an information system $A$, $a(x)$ is the value of tuple $x$ in attribute $a$, $a(y)$ is the value of tuple y in attribute $a$, and $|A|$ is the cardinality of a set $A$.

## 3 Classification Approach based on Rough Mereology

The proposed classification approach applied Rough mereology and rough inclusion approaches in order to classify input datasets into sets of generated rules in three phases; namely pre-processing, clustering, and voting by objects. For pre-processing phase, the input dataset was mapped into a normalized dataset of the range [0,1], where the values 0 and 1 represent the smallest the largest values in each dataset's attribute, respectively. For clustering phase, rough mereology and rough inclusion algorithms were used for classifying the resulted dataset into sets of granules with different radius and produce rough inclusion table that reflects similarity degree among parameters. Finally, for voting by objects phase, voting by training objects algorithm was applied to select the optimized granules of dataset and to produce the optimal similarity measurement. Algorithm 3 shows the detailed steps for the proposed rough mereology based classification approach.

The proposed rough mereology based classification approach used the theory of rough mereology and fuzzification to classify the dataset into sets of generated rules and works in three phases; namely pre-processing, clustering, and voting by objects phases, as shown in figure 1.

1. **Pre-processing phase:** maps the experienced dataset into a normalized dataset of the range [0,1], where the values 0 and 1 represent the smallest the largest values in each dataset's attribute, respectively.
2. **Clustering phase:** uses theory of rough mereology and rough inclusion technique to classify the dataset into sets of granules with different radius.
3. **Voting by objects phase:** applies voting by objects to select the optimized granules of dataset.

To describe the proposed rough mereology based classification approach, we considered the Abalone dataset in UCI [8] that consists of 4177 objects as an illustrative example. Table 1, clarifies a sample part of the Abalone dataset, which consists of eight attributes: Sex (S), Length (L), Diameter (D), Height (H), Whole weight (W), Shucked weight (SW), Viscera weight (V), and Shell weight (SH).

**Algorithm 1** Classification Approach based on Rough Mereology

```
Input: An information System table (IST)
Output: Rule granule set (RGS)
Initialize set of Rule sets SRS = ϕ
for Each radius from 0 to 1 step 0.1 do
    Initialize Rough Inclusion set RIS = ϕ
    for Each column in (IST) do
        Compute rough inclusion and put the values in RIS
    end for
    Initialize Rule generation set RGS = ϕ
    for Each row in RIS do
        Get the rule according to definition of rough inclusion, and put the rule in RGS
        Remove duplicates rules from RGS
        Output rule generation set RGS
    end for
    Store RGS in SRS
end for
Initialize ACCRateset = ϕ
for Each rules in SRS do
    compute Accuracy measure rate and put in ACCRateset as shown in equation (7)
end for
Initialize bestACC as the first element in set ACCRateset
for Each element in ACCRateset do
    if element ¿ bestACC then
        bestACCset=element
    end if
end for
for Each element in ACCRateset do
    if element = bestACC then
        Store element in bestACCset
    end if
end for
get the corresponding set of rules of bestACCset and put in RGS
apply voting by object equations as shown in to refine RGS
Exit
```

**Table 1** Sample part of the Abalone Dataset

| S | L | D | H | W | SW | V | SH |
|---|---|---|---|---|---|---|---|
| M | 0.455 | 0.365 | 0.095 | 0.514 | 0.2245 | 0.101 | 0.15 |
| M | 0.35 | 0.265 | 0.09 | 0.2255 | 0.0995 | 0.0485 | 0.07 |
| F | 0.53 | 0.42 | 0.135 | 0.677 | 0.2565 | 0.1415 | 0.21 |
| M | 0.44 | 0.365 | 0.125 | 0.516 | 0.2155 | 0.114 | 0.155 |
| I | 0.33 | 0.255 | 0.08 | 0.205 | 0.0895 | 0.0395 | 0.055 |
| I | 0.425 | 0.3 | 0.095 | 0.3515 | 0.141 | 0.0775 | 0.12 |
| F | 0.53 | 0.415 | 0.15 | 0.7775 | 0.237 | 0.1415 | 0.33 |
| F | 0.545 | 0.425 | 0.125 | 0.768 | 0.294 | 0.1495 | 0.26 |
| M | 0.475 | 0.37 | 0.125 | 0.5095 | 0.2165 | 0.1125 | 0.165 |
| F | 0.55 | 0.44 | 0.15 | 0.8945 | 0.3145 | 0.151 | 0.32 |

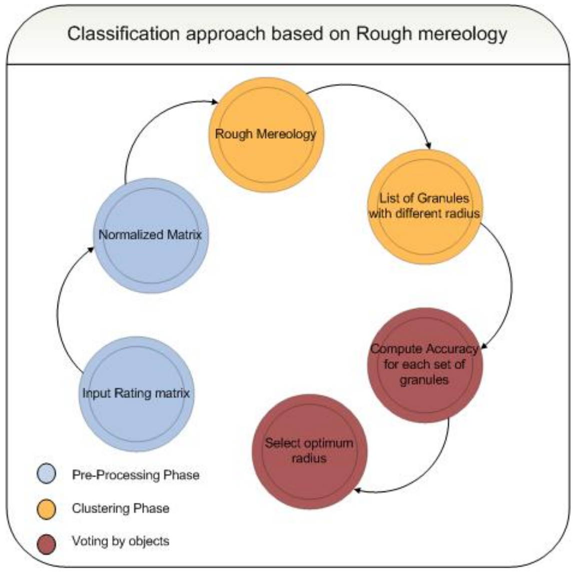


**Fig. 1** Phases of the proposed classification approach

## *3.1 Pre-processing phase*

Pre-processing phase takes dataset as a data input in the matrix form, called *rating matrix*, so that the proposed approach normalizes this matrix by finding the largest value and the smallest value in each attribute, which are called $max_a$ and $min_a$, respectively. The produced new matrix, *Normalized Matrix*, contains values in the range $[0,1]$ and is calculated as shown in equation (6):

$$NormalizedV_a = \frac{V_a - min_a}{max_a - min_a} \tag{6}$$

Table 2 shows the normalized matrix of Abalone Dataset presented in table 1.

**Table 2** Normalized matrix of Abalone Dataset

| L | D | H | W | SW | V | SH |
|---|---|---|---|---|---|---|
| 0.80 | 0.82 | 0.18 | 0.52 | 0.43 | 0.43 | 0.41 |
| 0.81 | 0.81 | 0.16 | 0.55 | 0.47 | 0.53 | 0.40 |
| 0.81 | 0.77 | 0.15 | 0.54 | 0.54 | 0.42 | 0.34 |
| 0.81 | 0.80 | 0.17 | 0.53 | 0.42 | 0.49 | 0.42 |
| 0.82 | 0.83 | 0.17 | 0.67 | 0.60 | 0.54 | 0.49 |
| 0.82 | 0.81 | 0.16 | 0.51 | 0.43 | 0.33 | 0.46 |
| 0.85 | 0.83 | 0.19 | 0.51 | 0.44 | 0.43 | 0.46 |
| 0.85 | 0.80 | 0.15 | 0.55 | 0.41 | 0.52 | 0.44 |
| 0.86 | 0.84 | 0.16 | 0.76 | 0.84 | 0.49 | 0.43 |
| 0.88 | 0.85 | 0.16 | 0.63 | 0.59 | 0.48 | 0.43 |

### 3.2 Clustering phase

Clustering phase received the normalized rating matrix and produced list of set of granules that describes rule generation according to different radius $r$ in granules. This phase consists of two stages; namely 1) rough mereology and 2) list of granules. For the rough mereology stage, the normalized rating matrix is considered as an input and the equations of rough mereology with radius $r$, previously described in section 2, were applied to that matrix. The output of this stage is rough inclusion table that represents similarity measure of each object in each attribute of the dataset. In the list of granules stage, a set of rough inclusion tables is computed by re-applying the first stage of rough mereology ten times with different radius of $r$ from 0.0 to 1.0 with step 0.1, as shown in table 3 and table 4, where table 3 describes the similarity measure at $r_0$ and table 4 describes the similarity measure at $r_2$.

**Table 3** Rough Inclusion of Abalone Dataset with $r = r_0$

| L | D | H | W | SW | V | SH |
|---|---|---|---|---|---|---|
| 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

**Table 4** Rough Inclusion of Abalone Dataset with $r = r_2$

| L | D | H | W | SW | V | SH |
|---|---|---|---|---|---|---|
| 1.000 | 0.997 | 0.022 | 0.677 | 0.576 | 0.577 | 0.585 |
| 0.997 | 1.000 | 0.025 | 0.680 | 0.579 | 0.580 | 0.588 |
| 0.022 | 0.025 | 1.000 | 0.345 | 0.446 | 0.445 | 0.436 |
| 0.677 | 0.680 | 0.345 | 1.000 | 0.896 | 0.896 | 0.902 |
| 0.576 | 0.579 | 0.446 | 0.896 | 1.000 | 0.915 | 0.900 |
| 0.577 | 0.580 | 0.445 | 0.896 | 0.915 | 1.000 | 0.914 |
| 0.585 | 0.588 | 0.436 | 0.902 | 0.900 | 0.914 | 1.000 |

### 3.3 Voting by objects

Voting by objects phase takes as an input the list of similarity measure tables for each radius resulted from the clustering phase and computes the accuracy rate for each radius as shown in equation (7). Then, the optimum radius $r_{opt}$ that represents the largest accuracy measure is selected. If more than one largest accuracy measure exist, then the proposed approach will generate set of largest accuracy measure. This phase is divided into two stages; namely 1) accuracy measure computation, which uses equation (7), where $T$ is the number of similar tuples and $N$ is the total number of tuples in each row, to compute the accuracy for each table representing the similarity measure at radius $r$ and 2) optimization stage, where the table containing the largest accuracy measure at radius $r$ is selected, so the radius $r$ called the *optimum radius* $r_{opt}$.

$$Accuracyrate = (\frac{T}{N}) * 100 \tag{7}$$

Results of the illustrative example tested in this article, considering the Abalone database, showing that the accuracy measure of table 3 is 87.5 and the accuracy measure of table 4 is 86.21. So, it could be concluded that when $r = r_0$ is the largest accuracy measure, $r_{opt}$ is $r_0$.

## 4 Experimental Results and Discussion

Simulation experiments have been conducted on 5 datasets; namely, Abalone Database (A), Car Evaluation database (CE), Nursery database (N), Pima Indian Diabetes database (PID), and Wisconsin Prognostic Breast database (WPE), shown in table 5 [8]. Table 6 represents the accuracy measure rate of all datasets with different values of $r$ in interval [0,1], where $r_0$ and $r_{10}$ represent the values of r = 0.0 and r =1.0, respectively, with step 0.1 for each increment of $r_i$, where i = 0, 1, ..., 10. The values of each tuple represents the accuracy measure for each radius $r$ in different datasets. For example, the accuracy measure for Abalone database at $r_5$ was shown to be less than the accuracy measure at $r_4$. So, as presented in table 6, the optimum radius in that case is the radius when $r = r_0$.

**Table 5** The Datasets

| Dataset | Number of records | Number of attributes |
|---|---|---|
| Abalone Data Database (A) | 4177 | 8 |
| Car Evaluation Database (CE) | 1782 | 6 |
| Nursery Database (N) | 12960 | 8 |
| Pima Indian Diabetes Database (PID) | 768 | 8 |
| Wisconsin Prognostic Breast Database (WPB) | 569 | 31 |

From table 3 and 4, we found out that the data of both tables differs when we increase the radius of granular $r_i$, where $r_i$ represents the radius of granules and the step of radius is $\frac{i}{10}$, where i = 0, 1, ..., 10. In the classification phase, we computed the inclusion, ten times as shown in table 6, where the radius changes from minimum value of each attribute to minimum value increased by step=0.1 for each iteration. So, we have ten inclusion tables with different granule's radius and has different data. The voting by objects phase is responsible to find the optimum radius, which gives more accurate rate for the generated rules. Initially, a part of the dataset is accustomed to generate rules with the proposed approach. Then, that accustomed part of the dataset tests the average accurate rate of rules. Simulation experimental results are shown in tables 6 that represents the average accurate measure rate for different values of radius and optimum radius and table 7 that represents a comparison between the proposed rough mereology based classification approach on the same dataset that experimented by researchers in different classifiers like ID3 and RGAGC. So, from

**Table 6** Accuracy measure rate with $r[0,1]$

| Radius \ Dataset | A | CE | N | PID | WPB |
|---|---|---|---|---|---|
| $r_0$ | 87.50 | 85.71 | 88.89 | 88.89 | 90.00 |
| $r_1$ | 87.48 | 49.26 | 55.56 | 40.80 | 26.09 |
| $r_2$ | 86.21 | 51.34 | 55.56 | 39.28 | 20.59 |
| $r_3$ | 58.11 | 78.15 | 51.85 | 41.08 | 20.59 |
| $r_4$ | 47.90 | 74.79 | 52.38 | 43.31 | 21.24 |
| $r_5$ | 32.47 | 83.42 | 52.38 | 50.09 | 43.95 |
| $r_6$ | 48.01 | 76.50 | 52.13 | 53.55 | 90.00 |
| $r_7$ | 59.00 | 0.00 | 13.79 | 66.71 | 90.00 |
| $r_8$ | 0.00 | 0.00 | 0.00 | 60.37 | 90.00 |
| $r_9$ | 0.00 | 0.00 | 0.00 | 0.00 | 90.00 |
| $r_{10}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

table 7 we conclude that the proposed approach outperformed the accuracy measure obtained by other approaches.

**Table 7** Accuracy measure rates for the proposed classification approach against RGAGC and ID3 classifiers

| Dataset | Accuracy measure rate | | |
|---|---|---|---|
| | The proposed approach | RGAGC | ID3 |
| Abalone Data Database | 87.50 | 33.7 | 6 |
| Car Evaluation Database | 85.71 | 66 | 63.2 |
| Nursery Database | 88.89 | 91.98 | 72.7 |
| Pima Indian Diabetes Database | 88.89 | 80.7 | 20.6 |
| Wisconsin Prognostic Breast Database | 90.00 | 0 | 0 |

# 5 Conclusions and Future Works

Granular Computing (GrC) aims to find a way to acquire knowledge for huge orderless very high dimensional perception information. Theory of rough mereology is a main classifier that is defined under granular computing. The primary contribution of the proposed classification approach, presented in this article, is that it returns rules from the granule space without considering the selection of attributes as though a lot of classic decision tree techniques and the ”false preserving” property of quotient space theory is applied. Experimental results depicted that the proposed approach is valid for many aspects as it has been tested for 5 different datasets. We observed that the optimal classifier obtained with the general radius $r = min$ gave better accuracy than the classifier on training objects with $r$ near to 1, which means that weighting heuristics slightly improves the quality of classification. Generally, experimental results using Irvine ML Repository datasets showed that the proposed rough mereology based classification approach outperformed both RGAGC and ID3 previously proposed algorithms. The performance measures used for the proposed rough mereology based classification approach are the same measures of the other approaches applied to the same datasets. For future work, we will consider to apply the proposed classification approach on datasets with text reviews for rule generation using opinion mining approach.

## References

1. Lin and Liau, C.-J.: Granular Computing and Rough Sets. Data Mining and Knowledge Discovery Handbook, Ed. IV, pp. 535–561, (2005).
2. ZHENG, H.Z., CHU, D.-H., and ZHAN, D.C.: Association Rule Algorithm Based on Bitmap and Granular Computing. Artificial Intelligence and Machine Learning (AIML) Journal, vol. 5, no. 3, pp. 51–54, (2005).
3. Guliato, D. and Santos, J.C.S.: Granular Computing and Rough Sets to Generate Fuzzy Rules. Proceedings of the International Conference on Image Analysis and Recognition (ICIAR 2009), Canada, pp. 317–326, (2009).
4. An, J., Wang, G., Wu, Y., and Gan, Q.: A Rule Generation Algorithm based on Granular Computing. Proceedings of the IEEE International Conference on Granular Computing, Beijing, China, vol. 1, pp. 102–107, (2005).
5. Veltkamp, R.C. and Hagedoorn, M.: Shape Similarity Measures, Properties, and Constructions. Proceedings of The 4th International Conference on Advances in Visual Information Systems (VISUAL 2000), Lyon, France, vol. 1929, pp. 467-476, Springer, (2000).
6. Lesniewski, S.: On the foundations of set theory. Topoi, vol. 2, pp. 7–52, (1982).
7. Polkowski, L. and Artiemjew, P.: Granular Computing in the Frame of Rough Mereology. A Case Study: Classification of Data into Decision Categories by Means of Granular Reflections of Data. International journal of intelligent systems, vol. 26, no. 6, pp. 555–571, (2011).
8. UC Irvine Machine Learning Repository Datasets. [Online:] http://www.ics.uci.edu./ $\sim$ mlearn/databases/