

ORIGINAL RESEARCH

Open Access



Reduction based on similarity and decision-making

Abd El Fattah A. El Atik

Correspondence:
aelatik5@yahoo.com
Department of Mathematics,
Faculty of Science, Tanta University,
Tanta, 31527 Egypt

Abstract

Reduction of attributes in an information system (IS) is a basic step for IS analysis. The original rough set model of Pawlak depends on an equivalence relation which is strongly constraints. This paper aims to use similarity classes and similarity degrees to obtain a reduction of IS and indicate an approach by using an example from biochemistry to get a quantitative structure activity relationship (QSAR). Moreover, signs of each attribute and degrees of memberships are computed to give a decision by using the degree of similarity. The suggested approach gives an increase in decision-making and decision accuracy.

Keywords: Protein energy, Rough set, Membership function, QSAR, Similarity, Decision-making

Mathematics Subject Classification (2010): 54C08, 54D10, 54C50, 54C60

Introduction and preliminaries

The rough set theory of Pawlak [1] is a mathematical tool to analyze the uncertain data. Pawlak classified data to a set of classes, and each class consists of objects [2]. This classification depends on an equivalence relation, and so all objects in the same class have the same importance, but this is not true in real life for all uncertain data. The notion of membership was studied in the view point of topological spaces in [3] and [4] via a general binary relation. This membership was extended from the original rough membership function [5]. Allam et al. [6] and Salama [7] gave a new approach for basic rough set concepts. They obtained some topological structures from rough sets. In [8], Yao et al. discussed the issue of classes overlapping in rough sets and they introduced a membership of an object to a class. The approximations is used to calculate the accuracy as in [9]. Several studies such as a chemical topology [10] have focused on the stability of structure bonds in proteins. Also, similarity on rough sets is one of these studies which interests in the protein energy of unfolding. Similarity based on rough sets was discussed in [11] and [12]. In 2017, Elshenawy et al. [13] studied the similarity of data via a weighted membership. Walczak and Massart [14] classified the data by giving range chosen. They used lower and upper approximations to calculate the accuracy. In this case, the intersection of all reducts gives the core of attributes. The topology on X generated by R is a

knowledge base for X , and indication of symptoms for a fixed disease can be seen through the topology [15]. Different notions of a membership function based on rough sets were introduced and studied in [16], [17], and [18]. A QSAR [19] constructs a mathematical model interconnected in its biological activity a set of structural descriptors of a set of chemical compounds. The main purpose of this paper is to study the reduction based on similarity. We give a comparison between a reduction by similarity and some other types of reduction with some different examples. An application on QSAR of AAs will be studied. We introduce an algorithm to reduce a membership function and illustrate it graphically. Some properties of relations and membership functions will be investigated. A new method is introduced to study the correlation between attributes and a decision through a similarity relation.

Definition 1 [1] *The IS or approximation space is a system (U, A, f) where U is the universe of finite set of objects and A is a set of attributes which is featured or variables. Each $a \in A$ defines an information function $f_a : U \rightarrow V_a$ where V_a is the set of all values of a , say the domain of attribute a .*

Definition 2 [5] *For every $B \subseteq A$, the indiscernibility relation on B , denoted by $Ind(B)$, is defined to be two objects $x_i, x_j \in U$ which are indiscernible by the sets B if $b(x_i) = b(x_j)$ for every $b \in B$. $Ind(B)$ is the smallest indiscernible groups of objects and so the equivalence class of $Ind(B)$ called elementary set in B . $[x_i]$ will denote to the equivalence class of object x_i in the relation $Ind(B)$.*

Definition 3 [5] *For every $B \subseteq A$, a membership function of an object $x_i \in U$ with respect to B is given by $\mu_B(x_i) = \frac{|[x_i] \cap B|}{|[x_i]|}$, where $[x_i]$ is the cardinality the class $[x_i]$, in other words, the number of objects contained in $[x_i]$.*

The IS is represented in Table 1 with a set of attributes $A = \{p, q, r, s\}$ and a set of objects $U = \{a, b, c, d, e\}$. We study the effect of attributes on decision, in other words, the correlation between attributes and the decision in the following statements.

Original Pawlak method

In Table 1, we compare between attributes of objects. There is no similarity between attributes of objects. So, we have the class of objects $C = \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}\}$ and two sets of decisions $D_1 = \{a, d, e\}$ and $D_2 = \{b, c\}$. The Pawlak membership is given by $\mu_A(x) = \frac{|[x] \cap A|}{|[x]|}$, for any object x and $A \subseteq U$. Therefore:

$$\mu_{D_1}(x) = \begin{cases} 1 & : x \in \{a, d, e\} \\ 0 & : x \in \{b, c\} \end{cases}$$

Table 1 Information system

| Objects | p | q | r | s | Decision |
|---------|-------|-------|-------|-------|----------|
| a | 0.23 | 0.31 | -0.55 | 254.2 | 8.5 |
| b | -0.48 | -0.60 | 0.51 | 303.6 | 8.2 |
| c | -0.61 | -0.77 | 1.20 | 287.9 | 8.2 |
| d | 0.45 | 1.54 | -1.40 | 282.9 | 8.5 |
| e | -0.11 | -0.22 | 0.29 | 335.0 | 8.5 |

Table 2 Objects with attribute coding

| Attributes | 1 | 2 | 3 | 4 |
|------------|-----------|-------------------|------------------|-----------|
| <i>a</i> | < -0.115 | [-0.115, 0.54) | [0.54, 1.195) | > 1.195 |
| <i>b</i> | < -0.18 | [-0.18, 0.63) | [0.63, 1.44) | > 1.440 |
| <i>c</i> | < -1.725 | [-1.725, -0.75) | [-0.75, 0.225) | > 0.225 |
| <i>d</i> | < 269.125 | [269.125, 313.35) | [313.35, 357.58) | > 357.580 |
| <i>e</i> | < 2.685 | [2.685, 3.708) | [3.708, 4.732) | > 4.732 |

$$\mu_{D_2}(x) = \begin{cases} 1 & : x \in \{b, c\} \\ 0 & : x \in \{a, d, e\} \end{cases}$$

This means any object correlates with the decision only by 0 and 1.

Pawlak method with coding

The IS in Table 1 can be coded by choosing intervals as in Table 2. The IS with coding is given by Table 3. The class of objects in Table 3 is $C = \{\{a\}, \{b, c\}, \{d\}, \{e\}\}$. Objects correlate with decisions D_1 and D_2 by:

$$\mu_{D_1}(x) = \begin{cases} 1 & : x \in \{a, d, e\} \\ 0 & : x \in \{b, c\} \end{cases},$$

$$\mu_{D_2}(x) = \begin{cases} \frac{1}{2} & : x \in \{b, c\} \\ 0 & : x \in \{a, d, e\} \end{cases}$$

This means that any object correlates with the decision only by 0, $\frac{1}{2}$ and 1.

Similarity without degree

The similarity matrix between objects in Table 3 is given by Table 4. A binary relation R on U is defined by xRy if and only if $\mu(x, y) \leq \frac{1}{4}$. Then, the class of objects is $C = \{U, \{a, b, d\}, \{a, c, d\}, \{a, d, e\}$. The correlation of objects with respect to decisions D_1 and D_2 is given by

$$\mu_{D_1}(a) = \frac{|U \cap \{a, b, d\} \cap \{a, c, d\} \cap \{a, d, e\}|}{|D_1|} = \frac{|a, d|}{|D_1|} = \frac{2}{3}, \mu_{D_1}(b) = 1, \mu_{D_1}(c) = 1, \mu_{D_1}(d) = \frac{2}{3}, \mu_{D_1}(e) = 1.$$

Also, $\mu_{D_2}(a) = 1, \mu_{D_2}(b) = \frac{3}{2}, \mu_{D_2}(c) = \frac{3}{2}, \mu_{D_2}(d) = 1$ and $\mu_{D_2}(e) = \frac{3}{2}$.

Uncertain QSAR information system

The basic form of Pawlak depends on equivalence relation. This expands the application circle for objects and decision rules. Because if we apply Pawlak rules for some ISs, it is not possible to have equal attributes for two different attributes. So as in the following problem, that we are giving a method of solving depends on a similarity relation.

Problem A modeling of the energy of unfolding of a protein (tryptophane synthase an alpha unit of the bacteriophage T_4 lysozyme), where 19 coded amino acids (AAs) were

Table 3 Coded information system

| Objects | <i>p</i> | <i>q</i> | <i>r</i> | <i>s</i> | Decision |
|----------|----------|----------|----------|----------|----------|
| <i>a</i> | 2 | 2 | 3 | 1 | 8.5 |
| <i>b</i> | 1 | 1 | 4 | 2 | 8.2 |
| <i>c</i> | 1 | 1 | 4 | 2 | 8.2 |
| <i>d</i> | 2 | 4 | 2 | 2 | 8.5 |
| <i>e</i> | 2 | 1 | 4 | 3 | 8.5 |

Table 4 Similarity matrix

| Objects | <i>a</i> | <i>b</i> | <i>c</i> | <i>d</i> | <i>e</i> |
|----------|---------------|---------------|---------------|---------------|---------------|
| <i>a</i> | $\frac{1}{4}$ | 0 | 0 | $\frac{1}{4}$ | $\frac{1}{4}$ |
| <i>b</i> | 0 | $\frac{1}{4}$ | 1 | $\frac{1}{4}$ | $\frac{1}{2}$ |
| <i>c</i> | 0 | 1 | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{2}$ |
| <i>d</i> | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ |
| <i>e</i> | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{4}$ | $\frac{1}{4}$ |

each introduced into position 49 [20]. The AAs are described in terms of seven attributes: $a_1 = \text{PIE}$ and $a_2 = \text{PIF}$ (two measures of the side chain lipophilicity), $a_3 = \text{DGR} = \Delta G$ of transfer from the protein interior with water, $a_4 = \text{SAC} = \text{surface area}$, $a_5 = \text{MR} = \text{molecular refractivity}$, $a_6 = \text{LAM} = \text{the side chain polarity}$, and $a_7 = \text{Vol} = \text{molecular volume}$. In [14], the authors used the form of Pawlak [1] to make decision rules. The IS of quantitative attributes $\{a_1, a_2, a_3, a_4, a_5, a_6, a_7\}$ and a decision attribute $\{d\}$ can be represented by Table 5. The condition attributes are coded into four qualitative terms, such as very low, low, high, and very high, whereas the decision attributes is coded into three qualitative terms, such as low, medium, and high. The qualitative terms of all attributes are coded by integer numbers. The problem is to illustrate that there are some objects that coded with medium energy of unfolding in [14] in respect of an attribute would be for the high energy of unfold.

Algorithm

Step 1: Construct a similarity matrix for each attribute *a* by $M_a = [w_{ij}]$ which will be 19×19 matrix, 7 is the number of attributes, $i \in \{1, 2, 3, \dots, 19\}$ will denote to the row of matrix, and $j \in \{1, 2, 3, \dots, 19\}$ will denote to the column of the matrix.

Step 2: Calculate the similarity degree of attributes through a definite relation, say *R*, on the our IS which will be denoted by QSAR IS = (*U*, *A*), where $U = \{x_1, x_2, x_3, x_4, x_5,$

Table 5 QSAR information system

| Objects | <i>a</i> ₁ | <i>a</i> ₂ | <i>a</i> ₃ | <i>a</i> ₄ | <i>a</i> ₅ | <i>a</i> ₆ | <i>a</i> ₇ | Decision |
|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|----------|
| <i>x</i> ₁ | 0.23 | 0.31 | −0.55 | 254.2 | 2.126 | −0.02 | 82.2 | 8.5 |
| <i>x</i> ₂ | −0.48 | −0.60 | 0.51 | 303.6 | 2.994 | −1.24 | 112.3 | 8.2 |
| <i>x</i> ₃ | −0.61 | −0.77 | 1.20 | 287.9 | 2.994 | −1.08 | 103.7 | 8.5 |
| <i>x</i> ₄ | 0.45 | 1.54 | −1.40 | 282.9 | 2.933 | −0.11 | 99.1 | 11.0 |
| <i>x</i> ₅ | −0.11 | −0.22 | 0.29 | 335.0 | 3.458 | −1.19 | 127.5 | 6.3 |
| <i>x</i> ₆ | −0.51 | −0.64 | 0.76 | 311.6 | 3.243 | −1.43 | 120.5 | 8.8 |
| <i>x</i> ₇ | 0.00 | 0.00 | 0.00 | 224.9 | 1.662 | 0.03 | 65.0 | 7.1 |
| <i>x</i> ₈ | 0.15 | 0.13 | −0.25 | 337.2 | 3.856 | −1.06 | 140.6 | 10.1 |
| <i>x</i> ₉ | 1.20 | 1.80 | −2.10 | 322.6 | 3.350 | 0.04 | 131.7 | 16.8 |
| <i>x</i> ₁₀ | 1.28 | 1.70 | −2.00 | 324.0 | 3.518 | 0.12 | 131.5 | 15.0 |
| <i>x</i> ₁₁ | −0.77 | −0.99 | 0.78 | 336.6 | 2.933 | −2.26 | 144.3 | 7.9 |
| <i>x</i> ₁₂ | 0.90 | 1.23 | −1.60 | 336.3 | 3.860 | −0.33 | 132.3 | 13.3 |
| <i>x</i> ₁₃ | 1.56 | 1.79 | −2.60 | 336.1 | 4.638 | −0.05 | 155.8 | 11.2 |
| <i>x</i> ₁₄ | 0.38 | 0.49 | −1.50 | 228.5 | 2.876 | −0.31 | 106.7 | 8.2 |
| <i>x</i> ₁₅ | 0.00 | −0.04 | 0.09 | 266.7 | 2.279 | −0.40 | 88.5 | 7.4 |
| <i>x</i> ₁₆ | 0.17 | 0.26 | −0.58 | 282.9 | 2.743 | −0.53 | 105.3 | 8.8 |
| <i>x</i> ₁₇ | 1.85 | 2.25 | −2.70 | 401.8 | 5.755 | −0.31 | 185.9 | 9.9 |
| <i>x</i> ₁₈ | 0.89 | 0.96 | −1.70 | 377.8 | 4.791 | −0.84 | 162.7 | 8.8 |
| <i>x</i> ₁₉ | 0.71 | 1.22 | −1.60 | 295.1 | 3.054 | −0.13 | 115.6 | 12.0 |

$x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}$ and $A = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7\}$, the set of attributes. So, $x_i R x_j$ if and only if x_i is similar to x_j . The degree of similarity or similarity measure between x_i and x_j is denoted by $\text{deg}_a(x_i, x_j) = d_{ij} = 1 - \frac{|a(x_i) - a(x_j)|}{|a_{\max} - a_{\min}|}$, where a_{\max} and a_{\min} denote the minimum and maximum values of attribute a , respectively. It is clear that R is reflexive and symmetric.

Step 3: Classify the data deduced from step 2 via $x_i R = \{x_j(d_{ij}) : x_i R x_j\}$. The value d will be chosen by an expert. Then, there are two cases $d_{ij} > d$ and $d_{ij} < d$.

Step 4: Define a membership function of every object for the similarity matrix in step 3 as follows: For any arbitrary element $x_k R = C_k \in U/R$ and for $y \in C_k$, the membership function of y with respect to any subset X of U , $\mu_X^{C_k}(y) = \frac{\sum_{x \in C_k \cap X} d_{ij}}{\sum_{x \in C_k} d_{ij}}$, where

$k \in \{1, 2, \dots, 19\}$. If $C_k \cap X = \phi$, then $\mu_X^{C_k}(y) = 0$.

Step 5: Present functions of the least, most extreme, and normal weighted participation for every y by:

$$\mu_X^w(y) = \min_{y \in C_k} (\mu_X^{C_k}(y))$$

$$\mu_X^w(y) = \max_{y \in C_k} (\mu_X^{C_k}(y))$$

$$\mu_X^w(y) = \text{avg}_{y \in C_k} (\mu_X^{C_k}(y))$$

Step 6: Choose a set $X \subset U$. Evaluate the rough membership μ for each object z by determining the $A = \bigcap_{z \in C_k} C_k$ and calculate $\mu = \frac{|A \cap X|}{|A|}$, for every object.

Step 7: Determine a maximum rough membership μ from the last column of each classification.

Now, we apply the algorithm on the problem of QSAR.

- (1) Similarity matrix for an attribute a_1 . Since $a_1(\max) = 1.85$ and $a_1(\min) = -0.77$, then, $a_1(\max) - a_1(\min) = 2.62$, then we have the matrix M_{x_1} . We have:

$$M_{a_1} = \begin{bmatrix} 1 & 0.7 & 0.7 & 1.0 & 0.9 & 0.7 & 0.9 & 1.0 & 0.9 & 0.6 & 0.6 & 0.7 & 0.5 & 0.9 & 0.9 & 1.0 & 0.4 & 0.7 & 0.8 \\ 0.7 & 1 & 0.9 & 0.6 & 0.9 & 1.0 & 0.8 & 0.8 & 0.4 & 0.3 & 0.9 & 0.9 & 0.5 & 0.2 & 0.7 & 0.8 & 0.7 & 0.5 & 0.5 \\ 0.7 & 0.9 & 1 & 0.6 & 0.8 & 0.9 & 0.8 & 0.7 & 0.3 & 0.3 & 0.9 & 0.4 & 0.2 & 0.6 & 0.8 & 0.7 & 0.1 & 0.4 & 0.5 \\ 1.0 & 0.6 & 0.6 & 1 & 0.8 & 0.6 & 0.8 & 0.9 & 0.7 & 0.7 & 0.5 & 0.8 & 0.6 & 0.9 & 0.8 & 0.9 & 0.5 & 0.8 & 0.8 \\ 0.9 & 0.9 & 0.8 & 0.8 & 1 & 0.8 & 0.9 & 0.9 & 0.5 & 0.5 & 0.7 & 0.6 & 0.4 & 0.8 & 0.9 & 0.9 & 0.2 & 0.6 & 0.4 \\ 0.7 & 1.0 & 0.9 & 0.6 & 0.8 & 1 & 0.8 & 0.7 & 0.3 & 0.3 & 0.9 & 0.5 & 0.2 & 0.7 & 0.8 & 0.7 & 0.1 & 0.5 & 0.5 \\ 0.9 & 0.8 & 0.8 & 0.8 & 0.9 & 0.8 & 1 & 0.9 & 0.5 & 0.5 & 0.7 & 0.7 & 0.4 & 0.8 & 1.0 & 0.9 & 0.3 & 0.7 & 0.7 \\ 1.0 & 0.8 & 0.7 & 0.9 & 0.9 & 0.7 & 0.9 & 1 & 0.6 & 0.6 & 0.6 & 0.7 & 0.5 & 0.9 & 0.9 & 1.0 & 0.3 & 0.7 & 0.8 \\ 0.9 & 0.4 & 0.3 & 0.7 & 0.5 & 0.3 & 0.5 & 0.6 & 1 & 0.9 & 0.2 & 0.9 & 0.9 & 0.7 & 0.5 & 0.6 & 0.7 & 0.9 & 0.8 \\ 0.6 & 0.3 & 0.3 & 0.7 & 0.5 & 0.3 & 0.5 & 0.6 & 0.9 & 1 & 0.2 & 0.8 & 0.9 & 0.7 & 0.5 & 0.6 & 0.8 & 0.8 & 0.8 \\ 0.6 & 0.9 & 0.9 & 0.5 & 0.7 & 0.9 & 0.7 & 0.6 & 0.2 & 0.2 & 1 & 0.4 & 0.1 & 0.6 & 0.4 & 0.6 & 0.0 & 0.4 & 0.4 \\ 0.7 & 0.9 & 0.4 & 0.8 & 0.6 & 0.5 & 0.7 & 0.7 & 0.9 & 0.8 & 0.7 & 1 & 0.7 & 0.8 & 0.7 & 0.7 & 0.6 & 1.0 & 0.9 \\ 0.5 & 0.5 & 0.2 & 0.6 & 0.4 & 0.2 & 0.4 & 0.5 & 0.9 & 0.9 & 0.1 & 0.7 & 1 & 0.5 & 0.4 & 0.5 & 0.9 & 0.7 & 0.7 \\ 0.9 & 0.2 & 0.6 & 0.9 & 0.8 & 0.7 & 0.8 & 0.9 & 0.7 & 0.7 & 0.6 & 0.8 & 0.5 & 1 & 0.8 & 0.9 & 0.4 & 0.8 & 0.9 \\ 0.9 & 0.7 & 0.8 & 0.8 & 0.9 & 0.8 & 1.0 & 0.9 & 0.5 & 0.5 & 0.7 & 0.7 & 0.4 & 0.8 & 1 & 0.9 & 0.3 & 0.7 & 0.7 \\ 1.0 & 0.8 & 0.7 & 0.9 & 0.9 & 0.7 & 0.9 & 1.0 & 0.6 & 0.6 & 0.6 & 0.7 & 0.5 & 0.9 & 0.9 & 1 & 0.3 & 0.7 & 0.8 \\ 0.4 & 0.7 & 0.1 & 0.5 & 0.2 & 0.1 & 0.3 & 0.3 & 0.7 & 0.8 & 0.0 & 0.6 & 0.9 & 0.4 & 0.3 & 0.3 & 1 & 0.6 & 0.6 \\ 0.7 & 0.5 & 0.4 & 0.8 & 0.6 & 0.5 & 0.7 & 0.7 & 0.9 & 0.8 & 0.4 & 1.0 & 0.7 & 0.8 & 0.7 & 0.7 & 0.6 & 1 & 0.9 \\ 0.8 & 0.5 & 0.5 & 0.8 & 0.7 & 0.5 & 0.7 & 0.8 & 0.8 & 0.8 & 0.4 & 0.9 & 0.7 & 0.9 & 0.7 & 0.8 & 0.6 & 0.9 & 1 \end{bmatrix}$$

(2) Take $d = 0.7$, we have the similarity classes.

- $C_1 = x_1R = \{x_1(1), x_4(1.0), x_5(0.9), x_7(0.9), x_8(1.0), x_9(0.9), x_{14}(0.9), x_{15}(0.9), x_{16}(1.0), x_{19}(0.8)\};$
- $C_2 = x_2R = \{x_2(1), x_3(0.9), x_5(0.9), x_6(1.0), x_7(0.8), x_8(0.8), x_{11}(0.9), x_{12}(0.9), x_{16}(0.8)\};$
- $C_3 = x_3R = \{x_2(0.9), x_3(1), x_5(0.8), x_6(0.9), x_7(0.8), x_{11}(0.9), x_{15}(0.8)\};$
- $C_4 = x_4R = \{x_1(1.0), x_4(1), x_5(0.8), x_7(0.8), x_8(0.9), x_{12}(0.8), x_{14}(0.9), x_{15}(0.8), x_{16}(0.9), x_{18}(0.8), x_{19}(0.8)\};$
- $C_5 = x_5R = \{x_1(0.9), x_2(0.9), x_3(0.8), x_4(0.8), x_5(1), x_6(0.8), x_7(0.9), x_8(0.9), x_{14}(0.8), x_{15}(0.9), x_{16}(0.9)\};$
- $C_6 = x_6R = \{x_2(1.0), x_3(0.9), x_5(0.8), x_6(1), x_7(0.8), x_{11}(0.9), x_{15}(0.8)\};$
- $C_7 = x_7R = \{x_1(0.9), x_2(0.8), x_3(0.8), x_4(0.8), x_5(0.9), x_6(0.8), x_7(1), x_8(0.9), x_{14}(0.8), x_{15}(1.0), x_{16}(0.9)\};$
- $C_8 = x_8R = \{x_1(1.0), x_2(0.8), x_4(0.9), x_5(0.9), x_7(0.9), x_8(1), x_{14}(0.9), x_{15}(0.9), x_{16}(1.0), x_{19}(0.8)\};$
- $C_9 = x_9R = \{x_1(0.9), x_9(1), x_{10}(0.9), x_{12}(0.9), x_{13}(0.9), x_{18}(0.9), x_{19}(0.8)\};$
- $C_{10} = x_{10}R = \{x_9(0.9), x_{10}(1), x_{12}(0.8), x_{13}(0.9), x_{17}(0.8), x_{18}(0.8), x_{19}(0.8)\};$
- $C_{11} = x_{11}R = \{x_2(0.9), x_3(0.9), x_6(0.9), x_{11}(1)\};$
- $C_{12} = x_{12}R = \{x_2(0.9), x_4(0.8), x_9(0.9), x_{10}(0.8), x_{12}(1), x_{14}(0.8), x_{18}(1.0), x_{19}(0.9)\};$
- $C_{13} = x_{13}R = \{x_9(0.9), x_{10}(0.9), x_{13}(1), x_{17}(0.9)\};$
- $C_{14} = x_{14}R = \{x_1(0.9), x_4(0.9), x_5(0.8), x_7(0.8), x_8(0.9), x_{12}(0.8), x_{14}(1), x_{15}(0.8), x_{16}(0.9), x_{18}(0.8), x_{19}(0.9)\};$
- $C_{15} = x_{15}R = \{x_1(0.9), x_3(0.8), x_4(0.8), x_5(0.9), x_6(0.8), x_7(1.0), x_8(0.9), x_{14}(0.8), x_{15}(1), x_{16}(0.9)\};$
- $C_{16} = x_{16}R = \{x_1(1.0), x_2(0.8), x_4(0.9), x_5(0.9), x_7(0.9), x_8(1.0), x_{14}(0.9), x_{15}(0.9), x_{16}(1), x_{19}(0.8)\};$
- $C_{17} = x_{17}R = \{x_{10}(0.8), x_{13}(0.9), x_{17}(1)\};$
- $C_{18} = x_{18}R = \{x_4(0.8), x_9(0.9), x_{10}(0.8), x_{12}(1.0), x_{14}(0.8), x_{18}(1), x_{19}(0.9)\};$
- $C_{19} = x_{19}R = \{x_1(0.8), x_4(0.8), x_8(0.8), x_9(0.8), x_{10}(0.8), x_{12}(0.9), x_{14}(0.9), x_{16}(0.8), x_{18}(0.9), x_{19}(1)\}.$

(3) We choose some various sets and compare the membership classification through the following two cases.

Case 1: Let $X = \{x_9, x_{10}, x_{12}\}$. Then, we have:

$$\begin{array}{lllll}
 \mu_X^{C_1}(x_1) = 0.10 & \mu_X^{C_2}(x_2) = 0.11 & \mu_X^{C_2}(x_3) = 0.11 & \mu_X^{C_1}(x_4) = 0.10 & \mu_X^{C_1}(x_5) = 0.10 \\
 \mu_X^{C_4}(x_1) = 0.08 & \mu_X^{C_3}(x_2) = 0.0 & \mu_X^{C_3}(x_3) = 0.0 & \mu_X^{C_4}(x_4) = 0.08 & \mu_X^{C_2}(x_5) = 0.11 \\
 \mu_X^{C_5}(x_1) = 0.0 & \mu_X^{C_5}(x_2) = 0.0 & \mu_X^{C_5}(x_3) = 0.0 & \mu_X^{C_5}(x_4) = 0.0 & \mu_X^{C_3}(x_5) = 0.0 \\
 \mu_X^{C_7}(x_1) = 0.0 & \mu_X^{C_6}(x_2) = 0.0 & \mu_X^{C_6}(x_3) = 0.0 & \mu_X^{C_7}(x_4) = 0.0 & \mu_X^{C_4}(x_5) = 0.08 \\
 \mu_X^{C_8}(x_1) = 0.0 & \mu_X^{C_7}(x_2) = 0.0 & \mu_X^{C_7}(x_3) = 0.0 & \mu_X^{C_8}(x_4) = 0.0 & \mu_X^{C_5}(x_5) = 0.0 \\
 \mu_X^{C_9}(x_1) = 0.44 & \mu_X^{C_8}(x_2) = 0.0 & \mu_X^{C_{11}}(x_3) = 0.0 & \mu_X^{C_{12}}(x_4) = 0.13 & \mu_X^{C_6}(x_5) = 0.38 \\
 \mu_X^{C_{14}}(x_1) = 0.08 & \mu_X^{C_{11}}(x_2) = 0.0 & \mu_X^{C_{15}}(x_3) = 0.0 & \mu_X^{C_{14}}(x_4) = 0.08 & \mu_X^{C_7}(x_5) = 0.0 \\
 \mu_X^{C_{15}}(x_1) = 0.0 & \mu_X^{C_{12}}(x_2) = 0.38 & & \mu_X^{C_{15}}(x_4) = 0.0 & \mu_X^{C_8}(x_5) = 0.0 \\
 \mu_X^{C_{16}}(x_1) = 0.0 & \mu_X^{C_{16}}(x_2) = 0.0 & & \mu_X^{C_{16}}(x_4) = 0.0 & \mu_X^{C_{14}}(x_5) = 0.08 \\
 \mu_X^{C_{19}}(x_1) = 0.30 & & & \mu_X^{C_{18}}(x_4) = 0.44 & \mu_X^{C_{15}}(x_5) = 0.0 \\
 & & & \mu_X^{C_{19}}(x_4) = 0.30 & \mu_X^{C_{16}}(x_5) = 0.0
 \end{array}$$

and

$$\begin{array}{lllll}
 \mu_X^{C_2}(x_6) = 0.11 & \mu_X^{C_1}(x_7) = 0.10 & \mu_X^{C_1}(x_8) = 0.10 & \mu_X^{C_1}(x_9) = 0.10 & \mu_X^{C_9}(x_{10}) = 0.44 \\
 \mu_X^{C_3}(x_6) = 0.0 & \mu_X^{C_2}(x_7) = 0.11 & \mu_X^{C_2}(x_8) = 0.11 & \mu_X^{C_9}(x_9) = 0.44 & \mu_X^{C_{10}}(x_{10}) = 0.45
 \end{array}$$

$$\begin{aligned}
 \mu_X^{C_5}(x_6) = 0.0 & \quad \mu_X^{C_3}(x_7) = 0.0 & \quad \mu_X^{C_4}(x_8) = 0.08 & \quad \mu_X^{C_{10}}(x_9) = 0.45 & \quad \mu_X^{C_{12}}(x_{10}) = 0.38 \\
 \mu_X^{C_6}(x_6) = 0.0 & \quad \mu_X^{C_4}(x_7) = 0.08 & \quad \mu_X^{C_5}(x_8) = 0.0 & \quad \mu_X^{C_{12}}(x_9) = 0.38 & \quad \mu_X^{C_{13}}(x_{10}) = 0.49 \\
 \mu_X^{C_7}(x_6) = 0.0 & \quad \mu_X^{C_5}(x_7) = 0.0 & \quad \mu_X^{C_7}(x_8) = 0.0 & \quad \mu_X^{C_{13}}(x_9) = 0.49 & \quad \mu_X^{C_{17}}(x_{10}) = 0.30 \\
 \mu_X^{C_{11}}(x_6) = 0.0 & \quad \mu_X^{C_6}(x_7) = 0.0 & \quad \mu_X^{C_8}(x_8) = 0.0 & \quad \mu_X^{C_{18}}(x_9) = 0.44 & \quad \mu_X^{C_{18}}(x_{10}) = 0.44 \\
 \mu_X^{C_{15}}(x_6) = 0.0 & \quad \mu_X^{C_7}(x_7) = 0.0 & \quad \mu_X^{C_{14}}(x_8) = 0.08 & \quad \mu_X^{C_{19}}(x_9) = 0.30 & \quad \mu_X^{C_{19}}(x_{10}) = 0.30 \\
 & \quad \mu_X^{C_8}(x_7) = 0.0 & \quad \mu_X^{C_{15}}(x_8) = 0.0 & & \\
 & \quad \mu_X^{C_{14}}(x_7) = 0.08 & \quad \mu_X^{C_{16}}(x_8) = 0.0 & & \\
 & \quad \mu_X^{C_{15}}(x_7) = 0.0 & \quad \mu_X^{C_{19}}(x_8) = 0.30 & & \\
 & \quad \mu_X^{C_{16}}(x_7) = 0.0 & & &
 \end{aligned}$$

and

$$\begin{aligned}
 \mu_X^{C_2}(x_{11}) = 0.11 & \quad \mu_X^{C_2}(x_{12}) = 0.11 & \quad \mu_X^{C_9}(x_{13}) = 0.44 & \quad \mu_X^{C_1}(x_{14}) = 0.10 & \quad \mu_X^{C_1}(x_{15}) = 0.10 \\
 \mu_X^{C_3}(x_{11}) = 0.0 & \quad \mu_X^{C_4}(x_{12}) = 0.08 & \quad \mu_X^{C_{10}}(x_{13}) = 0.45 & \quad \mu_X^{C_4}(x_{14}) = 0.08 & \quad \mu_X^{C_3}(x_{15}) = 0.0 \\
 \mu_X^{C_6}(x_{11}) = 0.0 & \quad \mu_X^{C_9}(x_{12}) = 0.44 & \quad \mu_X^{C_{13}}(x_{13}) = 0.49 & \quad \mu_X^{C_5}(x_{14}) = 0.0 & \quad \mu_X^{C_4}(x_{15}) = 0.08 \\
 \mu_X^{C_{11}}(x_{11}) = 0.0 & \quad \mu_X^{C_{10}}(x_{12}) = 0.45 & \quad \mu_X^{C_{17}}(x_{13}) = 0.30 & \quad \mu_X^{C_7}(x_{14}) = 0.0 & \quad \mu_X^{C_5}(x_{15}) = 0.0 \\
 & \quad \mu_X^{C_{12}}(x_{12}) = 0.38 & & \quad \mu_X^{C_8}(x_{14}) = 0.0 & \quad \mu_X^{C_6}(x_{15}) = 0.0 \\
 & \quad \mu_X^{C_{14}}(x_{12}) = 0.08 & & \quad \mu_X^{C_{12}}(x_{14}) = 0.38 & \quad \mu_X^{C_7}(x_{15}) = 0.0 \\
 & \quad \mu_X^{C_{18}}(x_{12}) = 0.44 & & \quad \mu_X^{C_{14}}(x_{14}) = 0.08 & \quad \mu_X^{C_8}(x_{15}) = 0.0 \\
 & \quad \mu_X^{C_{19}}(x_{12}) = 0.30 & & \quad \mu_X^{C_{15}}(x_{14}) = 0.0 & \quad \mu_X^{C_{14}}(x_{15}) = 0.08 \\
 & & & \quad \mu_X^{C_{16}}(x_{14}) = 0.0 & \quad \mu_X^{C_{15}}(x_{15}) = 0.0 \\
 & & & \quad \mu_X^{C_{18}}(x_{14}) = 0.44 & \quad \mu_X^{C_{16}}(x_{15}) = 0.0 \\
 & & & \quad \mu_X^{C_{19}}(x_{14}) = 0.30 &
 \end{aligned}$$

and

$$\begin{aligned}
 \mu_X^{C_1}(x_{16}) = 0.10 & \quad \mu_X^{C_{10}}(x_{17}) = 0.45 & \quad \mu_X^{C_4}(x_{18}) = 0.08 & \quad \mu_X^{C_1}(x_{19}) = 0.10 \\
 \mu_X^{C_2}(x_{16}) = 0.11 & \quad \mu_X^{C_{13}}(x_{17}) = 0.49 & \quad \mu_X^{C_9}(x_{18}) = 0.44 & \quad \mu_X^{C_4}(x_{19}) = 0.08 \\
 \mu_X^{C_4}(x_{16}) = 0.08 & \quad \mu_X^{C_{17}}(x_{17}) = 0.30 & \quad \mu_X^{C_{10}}(x_{18}) = 0.45 & \quad \mu_X^{C_8}(x_{19}) = 0.0 \\
 \mu_X^{C_5}(x_{16}) = 0.0 & & \quad \mu_X^{C_{12}}(x_{18}) = 0.38 & \quad \mu_X^{C_9}(x_{19}) = 0.44 \\
 \mu_X^{C_7}(x_{16}) = 0.0 & & \quad \mu_X^{C_{14}}(x_{18}) = 0.08 & \quad \mu_X^{C_{10}}(x_{19}) = 0.45 \\
 \mu_X^{C_8}(x_{16}) = 0.0 & & \quad \mu_X^{C_{18}}(x_{18}) = 0.44 & \quad \mu_X^{C_{12}}(x_{19}) = 0.38 \\
 \mu_X^{C_{14}}(x_{16}) = 0.08 & & \quad \mu_X^{C_{19}}(x_{18}) = 0.30 & \quad \mu_X^{C_{14}}(x_{19}) = 0.08 \\
 \mu_X^{C_{15}}(x_{16}) = 0.0 & & & \quad \mu_X^{C_{16}}(x_{19}) = 0.0 \\
 \mu_X^{C_{16}}(x_{16}) = 0.0 & & & \quad \mu_X^{C_{18}}(x_{19}) = 0.44 \\
 \mu_X^{C_{19}}(x_{16}) = 0.30 & & & \quad \mu_X^{C_{19}}(x_{19}) = 0.30
 \end{aligned}$$

Now, we evaluate the rough membership for each object which gives a preferable value for the data. The object belongs to more than one class, so it has three membership minimum, maximum, and average. This can be shown in Table 6 and Fig. 1.

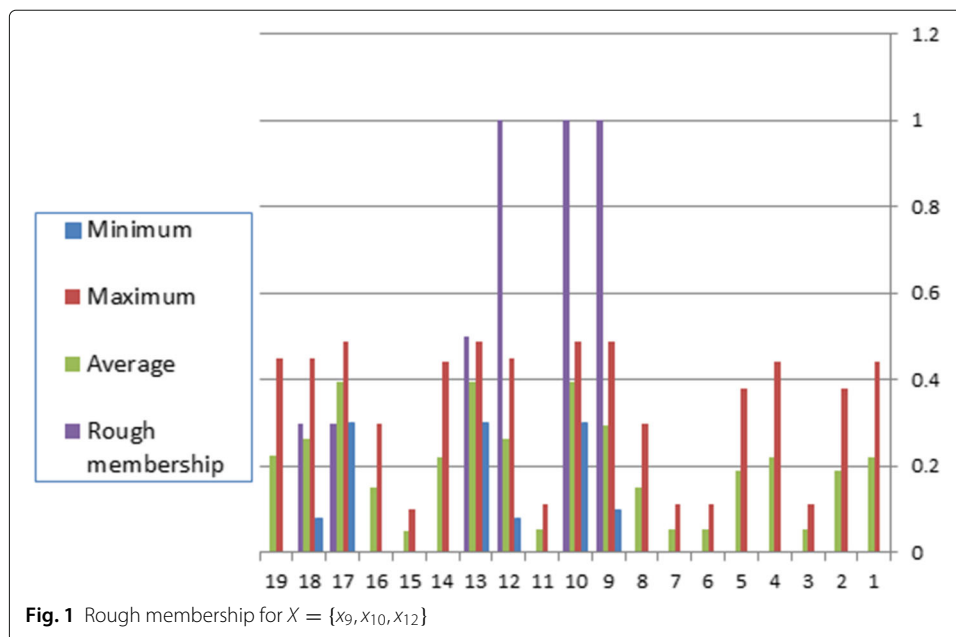
In [14], the decision rules are coded into three qualitative terms, such as low, medium, and high. X is the set of objects that protein has high energy of unfolding.

Table 6 Weighed rough membership for X

| Objects | Minimum | Maximum | Average | Rough membership μ |
|----------|---------|---------|---------|------------------------|
| x_1 | 0.0 | 0.44 | 0.22 | 0 |
| x_2 | 0.0 | 0.38 | 0.19 | 0 |
| x_3 | 0.0 | 0.11 | 0.055 | 0 |
| x_4 | 0.0 | 0.44 | 0.22 | 0 |
| x_5 | 0.0 | 0.38 | 0.19 | 0 |
| x_6 | 0.0 | 0.11 | 0.055 | 0 |
| x_7 | 0.0 | 0.11 | 0.055 | 0 |
| x_8 | 0.0 | 0.30 | 0.15 | 0 |
| x_9 | 0.10 | 0.49 | 0.295 | 1 |
| x_{10} | 0.30 | 0.49 | 0.395 | 1 |
| x_{11} | 0.0 | 0.11 | 0.055 | 0 |
| x_{12} | 0.08 | 0.45 | 0.265 | 1 |
| x_{13} | 0.30 | 0.49 | 0.395 | $\frac{1}{2}$ |
| x_{14} | 0.0 | 0.44 | 0.22 | 0 |
| x_{15} | 0.0 | 0.10 | 0.05 | 0 |
| x_{16} | 0.0 | 0.30 | 0.15 | 0 |
| x_{17} | 0.30 | 0.49 | 0.395 | $\frac{1}{3}$ |
| x_{18} | 0.08 | 0.45 | 0.265 | $\frac{1}{3}$ |
| x_{19} | 0.0 | 0.45 | 0.225 | 0 |

Case 2: Let $Y = \{x_4, x_8, x_{13}, x_{17}, x_{19}\}$. Then, we have:

$$\begin{aligned} \mu_Y^{C_1}(x_1) &= 0.30 & \mu_Y^{C_2}(x_2) &= 0.10 & \mu_Y^{C_3}(x_3) &= 0.10 & \mu_Y^{C_4}(x_4) &= 0.30 & \mu_Y^{C_5}(x_5) &= 0.30 \\ \mu_Y^{C_4}(x_1) &= 0.28 & \mu_Y^{C_3}(x_2) &= 0.0 & \mu_Y^{C_3}(x_3) &= 0.0 & \mu_Y^{C_4}(x_4) &= 0.28 & \mu_Y^{C_2}(x_5) &= 0.10 \\ \mu_Y^{C_5}(x_1) &= 0.18 & \mu_Y^{C_5}(x_2) &= 0.18 & \mu_Y^{C_5}(x_3) &= 0.18 & \mu_Y^{C_5}(x_4) &= 0.18 & \mu_Y^{C_3}(x_5) &= 0.0 \\ \mu_Y^{C_7}(x_1) &= 0.18 & \mu_Y^{C_6}(x_2) &= 0.0 & \mu_Y^{C_6}(x_3) &= 0.0 & \mu_Y^{C_7}(x_4) &= 0.18 & \mu_Y^{C_4}(x_5) &= 0.28 \\ \mu_Y^{C_8}(x_1) &= 0.30 & \mu_Y^{C_7}(x_2) &= 0.18 & \mu_Y^{C_7}(x_3) &= 0.18 & \mu_Y^{C_8}(x_4) &= 0.30 & \mu_Y^{C_5}(x_5) &= 0.18 \\ \mu_Y^{C_9}(x_1) &= 0.27 & \mu_Y^{C_8}(x_2) &= 0.30 & \mu_Y^{C_{11}}(x_3) &= 0.0 & \mu_Y^{C_{12}}(x_4) &= 0.24 & \mu_Y^{C_6}(x_5) &= 0.0 \\ \mu_Y^{C_{14}}(x_1) &= 0.28 & \mu_Y^{C_{11}}(x_2) &= 0.0 & \mu_Y^{C_{15}}(x_3) &= 0.19 & \mu_Y^{C_{14}}(x_4) &= 0.28 & \mu_Y^{C_7}(x_5) &= 0.18 \end{aligned}$$



$$\begin{aligned} \mu_{Y^{C_{15}}}^{C_{15}}(x_1) &= 0.19 & \mu_{Y^{C_{12}}}^{C_{12}}(x_2) &= 0.24 & \mu_{Y^{C_{15}}}^{C_{15}}(x_4) &= 0.19 & \mu_{Y^{C_8}}^{C_8}(x_5) &= 0.30 \\ \mu_{Y^{C_{16}}}^{C_{16}}(x_1) &= 0.30 & \mu_{Y^{C_{16}}}^{C_{16}}(x_2) &= 0.30 & \mu_{Y^{C_{16}}}^{C_{16}}(x_4) &= 0.30 & \mu_{Y^{C_{14}}}^{C_{14}}(x_5) &= 0.28 \\ \mu_{Y^{C_{19}}}^{C_{19}}(x_1) &= 0.31 & & & \mu_{Y^{C_{18}}}^{C_{18}}(x_4) &= 0.27 & \mu_{Y^{C_{15}}}^{C_{15}}(x_5) &= 0.19 \\ & & & & \mu_{Y^{C_{19}}}^{C_{19}}(x_4) &= 0.31 & \mu_{Y^{C_{16}}}^{C_{16}}(x_5) &= 0.30 \end{aligned}$$

and

$$\begin{aligned} \mu_{Y^{C_2}}^{C_2}(x_6) &= 0.10 & \mu_{Y^{C_1}}^{C_1}(x_7) &= 0.30 & \mu_{Y^{C_1}}^{C_1}(x_8) &= 0.30 & \mu_{Y^{C_1}}^{C_1}(x_9) &= 0.30 & \mu_{Y^{C_9}}^{C_9}(x_{10}) &= 0.27 \\ \mu_{Y^{C_3}}^{C_3}(x_6) &= 0.0 & \mu_{Y^{C_2}}^{C_2}(x_7) &= 0.10 & \mu_{Y^{C_2}}^{C_2}(x_8) &= 0.10 & \mu_{Y^{C_9}}^{C_9}(x_9) &= 0.27 & \mu_{Y^{C_{10}}}^{C_{10}}(x_{10}) &= 0.42 \\ \mu_{Y^{C_5}}^{C_5}(x_6) &= 0.18 & \mu_{Y^{C_3}}^{C_3}(x_7) &= 0.0 & \mu_{Y^{C_4}}^{C_4}(x_8) &= 0.28 & \mu_{Y^{C_{10}}}^{C_{10}}(x_9) &= 0.42 & \mu_{Y^{C_{12}}}^{C_{12}}(x_{10}) &= 0.24 \\ \mu_{Y^{C_6}}^{C_6}(x_6) &= 0.0 & \mu_{Y^{C_4}}^{C_4}(x_7) &= 0.28 & \mu_{Y^{C_5}}^{C_5}(x_8) &= 0.18 & \mu_{Y^{C_{12}}}^{C_{12}}(x_9) &= 0.24 & \mu_{Y^{C_{13}}}^{C_{13}}(x_{10}) &= 0.51 \\ \mu_{Y^{C_7}}^{C_7}(x_6) &= 0.18 & \mu_{Y^{C_5}}^{C_5}(x_7) &= 0.18 & \mu_{Y^{C_7}}^{C_7}(x_8) &= 0.18 & \mu_{Y^{C_{13}}}^{C_{13}}(x_9) &= 0.51 & \mu_{Y^{C_{17}}}^{C_{17}}(x_{10}) &= 0.70 \\ \mu_{Y^{C_{11}}}^{C_{11}}(x_6) &= 0.0 & \mu_{Y^{C_6}}^{C_6}(x_7) &= 0.0 & \mu_{Y^{C_8}}^{C_8}(x_8) &= 0.30 & \mu_{Y^{C_{18}}}^{C_{18}}(x_9) &= 0.27 & \mu_{Y^{C_{18}}}^{C_{18}}(x_{10}) &= 0.27 \\ \mu_{Y^{C_{15}}}^{C_{15}}(x_6) &= 0.19 & \mu_{Y^{C_7}}^{C_7}(x_7) &= 0.18 & \mu_{Y^{C_{14}}}^{C_{14}}(x_8) &= 0.28 & \mu_{Y^{C_{19}}}^{C_{19}}(x_9) &= 0.31 & \mu_{Y^{C_{19}}}^{C_{19}}(x_{10}) &= 0.31 \\ & & \mu_{Y^{C_8}}^{C_8}(x_7) &= 0.30 & \mu_{Y^{C_{15}}}^{C_{15}}(x_8) &= 0.19 & & & & \\ & & \mu_{Y^{C_{14}}}^{C_{14}}(x_7) &= 0.28 & \mu_{Y^{C_{16}}}^{C_{16}}(x_8) &= 0.30 & & & & \\ & & \mu_{Y^{C_{15}}}^{C_{15}}(x_7) &= 0.19 & \mu_{Y^{C_{19}}}^{C_{19}}(x_8) &= 0.31 & & & & \\ & & \mu_{Y^{C_{16}}}^{C_{16}}(x_7) &= 0.30 & & & & & & \end{aligned}$$

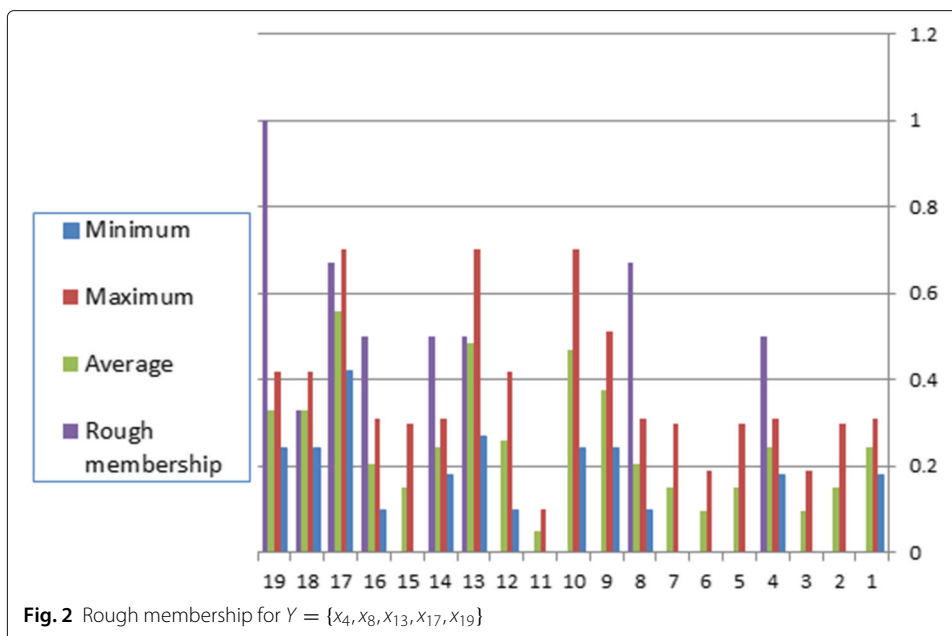
and

$$\begin{aligned} \mu_{Y^{C_2}}^{C_2}(x_{11}) &= 0.10 & \mu_{Y^{C_2}}^{C_2}(x_{12}) &= 0.10 & \mu_{Y^{C_9}}^{C_9}(x_{13}) &= 0.27 & \mu_{Y^{C_1}}^{C_1}(x_{14}) &= 0.30 & \mu_{Y^{C_1}}^{C_1}(x_{15}) &= 0.30 \\ \mu_{Y^{C_3}}^{C_3}(x_{11}) &= 0.0 & \mu_{Y^{C_4}}^{C_4}(x_{12}) &= 0.28 & \mu_{Y^{C_{10}}}^{C_{10}}(x_{13}) &= 0.42 & \mu_{Y^{C_4}}^{C_4}(x_{14}) &= 0.28 & \mu_{Y^{C_3}}^{C_3}(x_{15}) &= 0.0 \\ \mu_{Y^{C_6}}^{C_6}(x_{11}) &= 0.0 & \mu_{Y^{C_9}}^{C_9}(x_{12}) &= 0.27 & \mu_{Y^{C_{13}}}^{C_{13}}(x_{13}) &= 0.51 & \mu_{Y^{C_5}}^{C_5}(x_{14}) &= 0.18 & \mu_{Y^{C_4}}^{C_4}(x_{15}) &= 0.28 \\ \mu_{Y^{C_{11}}}^{C_{11}}(x_{11}) &= 0.0 & \mu_{Y^{C_{10}}}^{C_{10}}(x_{12}) &= 0.42 & \mu_{Y^{C_{17}}}^{C_{17}}(x_{13}) &= 0.70 & \mu_{Y^{C_7}}^{C_7}(x_{14}) &= 0.18 & \mu_{Y^{C_5}}^{C_5}(x_{15}) &= 0.18 \\ & & \mu_{Y^{C_{12}}}^{C_{12}}(x_{12}) &= 0.24 & & & \mu_{Y^{C_8}}^{C_8}(x_{14}) &= 0.30 & \mu_{Y^{C_6}}^{C_6}(x_{15}) &= 0.0 \\ & & \mu_{Y^{C_{14}}}^{C_{14}}(x_{12}) &= 0.28 & & & \mu_{Y^{C_{12}}}^{C_{12}}(x_{14}) &= 0.24 & \mu_{Y^{C_7}}^{C_7}(x_{15}) &= 0.18 \\ & & \mu_{Y^{C_{18}}}^{C_{18}}(x_{12}) &= 0.27 & & & \mu_{Y^{C_{14}}}^{C_{14}}(x_{14}) &= 0.28 & \mu_{Y^{C_8}}^{C_8}(x_{15}) &= 0.30 \\ & & \mu_{Y^{C_{19}}}^{C_{19}}(x_{12}) &= 0.31 & & & \mu_{Y^{C_{15}}}^{C_{15}}(x_{14}) &= 0.19 & \mu_{Y^{C_{14}}}^{C_{14}}(x_{15}) &= 0.28 \\ & & & & & & \mu_{Y^{C_{16}}}^{C_{16}}(x_{14}) &= 0.30 & \mu_{Y^{C_{15}}}^{C_{15}}(x_{15}) &= 0.19 \\ & & & & & & \mu_{Y^{C_{18}}}^{C_{18}}(x_{14}) &= 0.27 & \mu_{Y^{C_{16}}}^{C_{16}}(x_{15}) &= 0.30 \\ & & & & & & \mu_{Y^{C_{19}}}^{C_{19}}(x_{14}) &= 0.31 & & \end{aligned}$$

and

$$\begin{aligned} \mu_{Y^{C_1}}^{C_1}(x_{16}) &= 0.30 & \mu_{Y^{C_{10}}}^{C_{10}}(x_{17}) &= 0.42 & \mu_{Y^{C_4}}^{C_4}(x_{18}) &= 0.28 & \mu_{Y^{C_1}}^{C_1}(x_{19}) &= 0.30 \\ \mu_{Y^{C_2}}^{C_2}(x_{16}) &= 0.10 & \mu_{Y^{C_{13}}}^{C_{13}}(x_{17}) &= 0.51 & \mu_{Y^{C_9}}^{C_9}(x_{18}) &= 0.27 & \mu_{Y^{C_4}}^{C_4}(x_{19}) &= 0.28 \\ \mu_{Y^{C_4}}^{C_4}(x_{16}) &= 0.28 & \mu_{Y^{C_{17}}}^{C_{17}}(x_{17}) &= 0.70 & \mu_{Y^{C_{10}}}^{C_{10}}(x_{18}) &= 0.42 & \mu_{Y^{C_8}}^{C_8}(x_{19}) &= 0.30 \\ \mu_{Y^{C_5}}^{C_5}(x_{16}) &= 0.18 & & & \mu_{Y^{C_{12}}}^{C_{12}}(x_{18}) &= 0.24 & \mu_{Y^{C_9}}^{C_9}(x_{19}) &= 0.27 \\ \mu_{Y^{C_7}}^{C_7}(x_{16}) &= 0.18 & & & \mu_{Y^{C_{14}}}^{C_{14}}(x_{18}) &= 0.28 & \mu_{Y^{C_{10}}}^{C_{10}}(x_{19}) &= 0.42 \\ \mu_{Y^{C_8}}^{C_8}(x_{16}) &= 0.30 & & & \mu_{Y^{C_{18}}}^{C_{18}}(x_{18}) &= 0.27 & \mu_{Y^{C_{12}}}^{C_{12}}(x_{19}) &= 0.24 \\ \mu_{Y^{C_{14}}}^{C_{14}}(x_{16}) &= 0.28 & & & \mu_{Y^{C_{19}}}^{C_{19}}(x_{18}) &= 0.31 & \mu_{Y^{C_{14}}}^{C_{14}}(x_{19}) &= 0.28 \\ \mu_{Y^{C_{15}}}^{C_{15}}(x_{16}) &= 0.19 & & & & & \mu_{Y^{C_{16}}}^{C_{16}}(x_{19}) &= 0.30 \\ \mu_{Y^{C_{16}}}^{C_{16}}(x_{16}) &= 0.30 & & & & & \mu_{Y^{C_{18}}}^{C_{18}}(x_{19}) &= 0.27 \\ \mu_{Y^{C_{19}}}^{C_{19}}(x_{16}) &= 0.31 & & & & & \mu_{Y^{C_{19}}}^{C_{19}}(x_{19}) &= 0.31 \end{aligned}$$

Now, we evaluate the rough membership for each object which gives a preferable value for the data. This value will be evaluated via minimum, maximum, average, and weighed membership. This can be shown in Table 6 and Fig. 2.



Y is the set of objects that protein has a medium energy of unfolding. From Table 7, one can show that the objects x_9, x_{10}, x_{12} have a rough membership value 1; this means that a protein has high energy of unfolding. The object x_{19} in Pawlak reduction had a medium energy of unfolding [14], while from our procedure, x_{19} has a high energy of unfolding. In the same manner, we can take a set Z of a protein, which has a low energy of unfolding. Through the correlation between objects and decision, there is at least one object in Z which has the medium or the high energy of unfolding. Therefore, the significance of each

Table 7 Weighed rough membership for Y

| Objects | Minimum | Maximum | Average | Rough membership μ |
|----------|---------|---------|---------|------------------------|
| x_1 | 0.18 | 0.31 | 0.245 | 0 |
| x_2 | 0.0 | 0.30 | 0.15 | 0 |
| x_3 | 0.0 | 0.19 | 0.095 | 0 |
| x_4 | 0.18 | 0.31 | 0.245 | $\frac{1}{2}$ |
| x_5 | 0.0 | 0.30 | 0.15 | 0 |
| x_6 | 0.0 | 0.19 | 0.095 | 0 |
| x_7 | 0.0 | 0.30 | 0.15 | 0 |
| x_8 | 0.10 | 0.31 | 0.205 | $\frac{2}{3}$ |
| x_9 | 0.24 | 0.51 | 0.375 | 0 |
| x_{10} | 0.24 | 0.71 | 0.47 | 0 |
| x_{11} | 0.0 | 0.10 | 0.05 | 0 |
| x_{12} | 0.10 | 0.42 | 0.26 | 0 |
| x_{13} | 0.27 | 0.70 | 0.485 | $\frac{1}{2}$ |
| x_{14} | 0.18 | 0.31 | 0.245 | $\frac{1}{2}$ |
| x_{15} | 0.0 | 0.30 | 0.15 | 0 |
| x_{16} | 0.10 | 0.31 | 0.205 | $\frac{1}{2}$ |
| x_{17} | 0.42 | 0.70 | 0.56 | $\frac{2}{3}$ |
| x_{18} | 0.24 | 0.42 | 0.33 | $\frac{1}{3}$ |
| x_{19} | 0.24 | 0.42 | 0.33 | 1 |

attribute and degrees of memberships is more precise from Pawlak’s reduction.

We can evaluate analogously the quantitative for attributes $\{a_2, a_3, a_4, a_5, a_6, a_7\}$.

Some properties on a similarity relation

The given algorithm depends on a similarity matrix, general binary relation, and rough membership. So, we study some properties of these notions.

Proposition 1 *If R_1 and R_2 are two different relations, then the degree of similarity is the same.*

Proof A similarity measure between x_i and x_j is given by $\text{deg}_a(x_i, x_j) = 1 - \frac{|a(x_i) - a(x_j)|}{|a_{\max} - a_{\min}|}$. Since each of R_1 and R_2 depends on d , then $|x_i R_1| = \sum_{x_k \in C_k} d_{ij} = |x_i R_2|$. Therefore, the degree of similarity is the same. □

Proposition 2 *If $R_1 \subseteq R_2$, then $\text{deg}_{R_1}(x_i, x_j) \leq \text{deg}_{R_2}(x_i, x_j)$, where deg is the degree of similarity relation.*

Proof Since $R_1 \subseteq R_2$, then $|x_i R_1| = \sum_{x_k \in C_k} d_{ij} \leq \sum_{x_k \in C_k} d'_{ij} = |x_i R_2|$, where d_{ij} and d'_{ij} are the similarity degrees with respect to R_1 and R_2 , respectively. □

Proposition 3 *For an IS (U, A) , where U is the set of objects and A is the sets of attributes. Then, $\mu_X^{C_k}(x) = \mu_X^{C_k}(y)$, for every two different objects $x, y \in A$, every class C_k , and every nonempty set $X \subseteq U$.*

Proof Directly from $\mu_X^{C_k}(x) = \frac{\sum_{x \in C_k \cap X} d_{ij}}{\sum_{x \in C_k} d_{ij}} = \mu_X^{C_k}(y)$ □

Remark 1 *For an attribute $a \in A$ in an IS (U, A) . If $\mu_X^{C_k}(x) = \mu_X^{C_m}(x)$, $k \neq m$, it is not necessary that $C_k = C_m$. This is obvious in the problem of our study.*

Conclusion and discussion

Chemical data sets have been analyzed using similarity relations. The results are more precise in comparison with the original rough set theory. The description of objects is different from the study in [14]. For example, the decision concerning element x_{19} in our study has the high energy of unfolding, of which in [14], the energy was the medium of unfolding. This opened the way for applying similarity models in IS which give discrete structure and coincide with the classical case. The model of QSAR of similarity reduction can be applied to a finite set of objects. The approach used here can be applied in any IS with quantitative or qualitative data. Consequently, they are very significant in decision-making [21–24]. The introduced techniques are very useful in application because they open a way for more topological applications from real-life problems.

Abbreviations

IS: Information system; QSAR: Quantitative structure activity relationship; AAs: Amino acids

Acknowledgements

The author would like to thank the anonymous referees for their helpful comments that improve the presentation of this paper.

Authors' contributions

All authors jointly worked on the results, and they read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Competing interests

The author declares that there are no competing interests.

Received: 10 December 2019 Accepted: 27 March 2020

Published online: 13 May 2020

References

1. Pawlak, Z.: Rough sets. *Int. J. Inf. Comput. Sci.* **11**, 341–356 (1982). <https://doi.org/10.1007/BF01001956>
2. Pawlak, Z.: *Rough sets, Theoretical Aspects of Reasoning About Data*. Springer (1991)
3. Lashin, E. F., Kozae, A. M., Abo Khadra, A. A., Medhat, T.: Rough set theory for topological spaces. *Int. J. Approx. Reason.* **40**, 1–2, 35–43 (2005). <https://doi.org/10.1016/j.ijar.2004.11.007>
4. Salama, A. S., El-Barbary, O. G.: Topological approach to retrieve missing values in incomplete information systems. *J. Egypt. Math. Soc.* **25**, 419–423 (2017). <https://doi.org/10.1016/j.joems.2017.07.004>
5. Pawlak, Z., Skowron, A.: *Advances in the Dempster-Shafer theory of evidence. Chapter Rough membership functions*, 251–271. Wiley, New York (1994)
6. Allam, A. A., Bakeir, M. Y., Abo-Tabl, E. A.: *New approach for basic rough set concepts, Rough, Sets, Fuzzy Sets, Data Mining, and Granular Computing, Lecture Notes in Artificial Intelligence 3641* (Slezak, D., Wang, G., Szczuka, M., Dntsch, I., Yao, Y., eds.) Springer, Berlin (2005)
7. Salama, A. S.: Some topological properties of rough sets with tools for data mining, *IJCSI. Int. J. Comput. Sci.* **8**(2), 588–595 (2011)
8. Yao, Y. Y., Zhang, J. P.: Interpreting fuzzy membership functions in the theory of rough sets. In: *Rough Sets and Current Trends in Computing, Second International Conference*, pp. 82–89. RSTC 2000 Banff, Canada, (2000)
9. Yu, Z., Wang, D.: Accuracy of approximation operators during covering evolutions. *Int. J. Approx. Reason.* **117**, 1–14 (2020). <https://doi.org/10.1016/j.ijar.2019.10.012>
10. Wang, X. W., Zhang, W. B.: Chemical topology and complexity of protein architectures. *Trends Biochem. Sci.* **43**(10), 806–817 (2018). <https://doi.org/10.1016/j.tibs.2018.07.001>
11. Stepaniuk, J.: Similarity based rough sets and learning. In: *Proc. of 5th European Congress on Intelligent Techniques & Soft Computing*, pp. 1634–1638, (1997)
12. Abo-Tabl, E. A.: Rough sets and topological spaces based on similarity. *Int. J. Mach. Learn. and Cyber.* **4**, 451–458 (2013)
13. Elshenawy, A., El-Sayed, M. K., Elsodany, M.: Weighted membership based on matrix relation. *Math. Meth. Appl. Sci.*, 1–9 (2017). <https://doi.org/10.1002/mma.4631>
14. Walczak, B., Massart, D. L.: Tutorial rough set theory. *Chemometr. Intell. Lab. Syst.* **47**, 1–16 (1999)
15. Kozae, A. M.: On topology expansions by ideals and applications. *Chaos, Solitons and Fractals.* **13**, 55–60 (2002). [https://doi.org/10.1016/S0960-0779\(00\)00224-1](https://doi.org/10.1016/S0960-0779(00)00224-1)
16. Chakraborty, M. K.: Membership function based rough set. *Int. J. Approx. Reason.* **55**, 402–411 (2014). <https://doi.org/10.1016/j.ijar.2013.10.009>
17. Pawlak, Z., Skowron, A.: *Rough membership function: a tool for reasoning with uncertainty, Algebraic Methods in Logic and in Computer Science Banach Center Publications*, 28, Institute of Mathematics Polish Academy of Sciences Warszawa. **28**(1), 135–150 (1993)
18. El-Sayed, M. K.: Similarity based membership of elements to uncertain concept in information system. *Int. Sch. Sci. Res. Innov.* **12**(3), 58–61 (2018)
19. Tropsha, A.: *Fundamentals of QSAR modeling: basic concepts and applications, Fundamentals of Using QSAR Models and Read-across Techniques in Predictive Toxicology*, University of North Carolina. Chapel Hill, USA (2016)
20. El Tayar, N. E., Tsai, R.-S., Carrupt, P.-A., Testa, B. J. *Chem. Soc. Perkin Trans.* **2**, 79–84 (1992)
21. Jiang, H., Zhan, J., Chen, D.: Covering based variable precision $(\mathcal{I}, \mathcal{T})$ -fuzzy rough sets with applications to multi-attribute decision-making. *IEEE Trans. Fuzzy Syst.* (2018). <https://doi.org/10.1109/TFUZZ.2018.2883023>
22. Zhan, J., Sun, B., Alcantud, J. C. R.: Covering based multigranulation $(\mathcal{I}, \mathcal{T})$ -fuzzy rough set models and applications in multi attribute group decision-making. *Inf. Sci.* **476**, 290–318 (2019). <https://doi.org/10.1016/j.ijar.2013.05.004>
23. Zhang, L., Zhan, J., Xu, Z.: Covering-based generalized IF rough sets with applications to multi-attribute decision-making. *Inf. Sci.* **478**, 275–302 (2019). <https://doi.org/10.1016/j.ins.2018.11.033>
24. Alharbi, N., Aydi, H., Özel, C., Topal, S.: Rough topologies on classical and based covering rough sets with applications in making decisions on patients with some medical problems Preprint (2020). <https://www.researchgate.net/publication/338969347>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.