World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:12, No:3, 2018

# Similarity Based Membership of Elements to Uncertain Concept in Information System

M. Kamel El-Sayed

***Abstract*****—**The process of determining the degree of membership for an element to an uncertain concept has been found in many ways, using equivalence and symmetry relations in information systems. In the case of similarity, these methods did not take into account the degree of symmetry between elements. In this paper, we use a new definition for finding the membership based on the degree of symmetry. We provide an example to clarify the suggested methods and compare it with previous methods. This method opens the door to more accurate decisions in information systems.

***Keywords*****—**Information system, uncertain concept, membership function, similarity relation, degree of similarity.

International Science Index, Mathematical and Computational Sciences Vol:12, No:3, 2018 waset.org/Publication/10008697

## I. Introduction

THE present era is characterized by the ability of computers to collect and store data on any issue that we are considering for decision. The most difficult issues in research and analysis are issues of uncertainty and incomplete information, scientists have presented theories to deal with these subjects, the most important of which are the fuzzy set theory that has been proposed in [5], intuitionistic fuzzy set by [4] and rough set theory was introduced by [9]. The first theories require membership in [10] and non-membership functions based on experience. In rough set theory, information systems are used to find a membership function that has been based on equivalence relation [1]. This limits the scope of their application. Therefore, general models of the theory were introduced using similarity and dissimilarity relationships and thus created a membership function. Information systems are a general mathematical concept for analyzing uncertain data.

Studying and analyzing uncertain data in information systems plays an important role in real life areas and this approach needs to determine the degree of membership for an element to an uncertain concept. Pawlak used the equivalence relation [10], symmetry relations, topological relations were used in [3], [6], [8], we have observed that all these methods do not take into account the degree of symmetry. The aim of this paper is to present the methods of finding the membership function and to introduce a definition and construct examples and compare it with the pervious similar work. The rest of the paper is organized as follows: the second part presents the basic definitions of information system, the third part contains our definition, characteristics, the part four provides an application example, and the last part contains comparison with previous methods.

M. Kamel El-Sayed is with the Higher Institute of Engineering and Technology, Department of Basic Science, Kafr El-Sheikh, Egypt (e-mail: dr.mkamel123@gmail.com).

## II. Preliminaries

Rough set theory which was proposed by Pawlak [9], is a mathematical tool to deal with uncertainty and incomplete information. One of the considerable applications of rough set theory is the approximation and membership. In fact, the classical rough sets are difficult to handle well with a lot of data. For this, the relation of similarity rough sets [6], [7] has been applied.

**Definition 1** [5]: The information system (Information table) is a pair $IS=(U,A)$, where $U=\{x_1,x_2,...,\ x_n\}$ is a non empty finite set of objects called the universe and $A=\{a_1,a_2,...,a_n\}$ is a non-empty finite set of attributes such as a $U\to v_a$, $a\in A$. The set $v_a$ is called the value set of *a*. The set of attributes can be divided in two subsets $C\subset A$ and $D=A-C$, respectively the conditional set of attributes and the decision (or class) attribute(s). The following definition is a measure for dissimilarity objects based on the information table.

**Definition 2** [12]: *Dissimilarity measures:* Let the two objects be denoted as *n* attributes described by *A*, *B*. The measure of the dissimilarity between these two objects is defined as mismatches for them

$$d(A,B)=\sum_{i=1}^{m}\delta(a_i,b_i),\ \text{where}\ \delta(a_i,b_i)=\begin{vmatrix}0(a_i=b_i)\\1(a_i\neq b_i)\end{vmatrix}$$

One of the modern ways in artificial intelligence is to use the separation of elements near a certain element in the inferences about this element and all elements close to it. This method is applied for finding approximations of concepts, computing the accuracy of decisions and constructing membership functions to an uncertain concept. So, the case of equivalence relation can be obtained.

**Definition 3** [11]: Let $A=(U,A)$ be an information system and let $\phi\neq X\subseteq U$. The rough *A*-membership function of the set *X* (or rm-function, for short), denoted by $\mu_X^A$, is defined as:

$$\mu_X^A(x)=\frac{|(x)_A\cap X|}{|(x)_A|}\ \text{for}\ x\in U.$$

The above definition illustrated in Fig. 1.

International Science Index, Mathematical and Computational Sciences Vol:12, No:3, 2018 waset.org/Publication/10008697

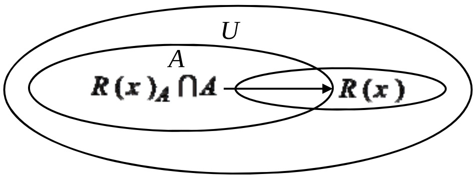


Fig. 1 Information system

## III. Similarity Based Membership.

The purpose of this article is to construct a matrix which represents similarity or dissimilarity relation and use the similarity degrees to construct a membership function. If $(u, R)$ is an $IsR$ is a similarity relation $u$ the similarity matrix is constant as follows:

$$M = a_{ij}, \qquad i, j = \{1, 2, ..., n\},$$

where the value of $a_{i,j}$ is denoted by $\mu(x_i, x_j)$.

If $(u, R)$ is an $Is$, $A C u$, $X \in$ the degree of member-ship of $x$ to $A$ denoted by $\mu_A(x)$ is given by

$$\mu_A(x) = \frac{\sum_{x_i \in A \cap R(x)} \{\mu(x, x_i), x \neq x_i\}}{\sum_{x_i \in R(x)} \mu(x, x_i), x \neq x_i}.$$

**Remark 1:** If $R(x) = \{y\}$ and $A \cap R(x) = \{y\}$, then

$$\mu_A(x) = \begin{cases} 1 & \text{if } R \text{ is similarity} \\ 0 & \text{if } R \text{ is dissimilarity.} \end{cases}$$

**Example 1***:* The process begins with the definition of attribute for each class. During the test phase in the power plant for the prototype, the pre-set was adjusted. Eight examples [2] are shown in Table I with the attributes: frequency, amplitude, (harmonic distortion level) THD, and distortion. Normal, warning, and danger are the outputs.

TABLE I
Test Phase in the Power Plant

| Ex. | Freq. | Amp. | TDH | Dis. | Out. |
|---|---|---|---|---|---|
| $x_1$ | Low | Nor. | Nor. | Nor. | Nor. |
| $x_2$ | Low | Med. | Med. | Nor. | Nor. |
| $x_3$ | Low | Med. | Nor. | High | Nor. |
| $x_4$ | Med. | Med. | Nor. | Med. | War. |
| $x_5$ | Med. | Med. | Nor. | High | War. |
| $x_6$ | Med. | High | Nor. | High | Dan. |
| $x_7$ | High | High | Med. | Med. | Dan. |
| $x_8$ | Med. | High | Nor. | Med. | Dan. |

Using Definition 2 we get the following dissimilarity matrix in Table II.

TABLE II
The Dissimilarity Matrix

| Ex. | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ |
|---|---|---|---|---|---|---|---|---|
| $x_1$ | 0 | 2 | 2 | 4 | 4 | 4 | 5 | 4 |
| $x_2$ | 2 | 0 | 2 | 4 | 4 | 5 | 4 | 4 |
| $x_3$ | 2 | 2 | 0 | 3 | 2 | 3 | 5 | 4 |
| $x_4$ | 4 | 4 | 3 | 0 | 1 | 3 | 4 | 2 |
| $x_5$ | 4 | 4 | 2 | 1 | 0 | 2 | 5 | 3 |
| $x_6$ | 4 | 5 | 3 | 3 | 2 | 0 | 3 | 1 |
| $x_7$ | 5 | 4 | 5 | 4 | 5 | 3 | 0 | 2 |
| $x_8$ | 4 | 4 | 4 | 2 | 3 | 1 | 2 | 0 |

Using the dissimilarity matrix, we get $R(x) = \{y : v(x, y) \leq \lambda\}$. We choose classes in more than one way according to the value of $\lambda$ as follows.

From Table II, in case where the output is Normal $A = \{x_1, x_2, x_3\}$

$$\mu(x) = \frac{A \cap R(x)}{R(x)}.$$

*Case I.* for $\lambda \leq 4$

$$R(x_1) = \{x_1, x_2, x_3, x_4, x_5, x_6, x_8\},$$
$$A \cap R(x_1) = \{x_1, x_2, x_3\}$$
$$\mu_A(x_1) = \frac{0+2+2}{0+2+2+4+4+4+4} = 0.25$$
$$R(x_2) = \{x_1, x_2, x_3, x_4, x_5, x_7, x_8\},$$
$$A \cap R(x_2) = \{x_1, x_2, x_3\}$$
$$\mu_A(x_2) = \frac{2+0+2}{2+0+2+4+4+4+4} = 0.25$$
$$R(x_3) = \{x_1, x_2, x_3, x_4, x_5, x_6, x_8\},$$
$$A \cap R(x_3) = \{x_1, x_2, x_3\}$$
$$\mu_A(x_3) = \frac{2+2+0}{2+2+0+3+2+3+4} = 0.25$$
$$R(x_4) = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\},$$
$$A \cap R(x_4) = \{x_1, x_2, x_3\}$$
$$\mu_A(x_4) = \frac{4+4+3}{4+4+3+0+1+3+4+2} = 0.52$$
$$R(x_5) = \{x_1, x_2, x_3, x_4, x_5, x_6, x_8\},$$
$$A \cap R(x_5) = \{x_1, x_2, x_3\}$$
$$\mu_A(x_5) = \frac{4+4+2}{4+4+2+1+0+2+3} = 0.625$$
$$R(x_6) = \{x_1, x_3, x_4, x_5, x_6, x_7, x_8\},$$
$$A \cap R(x_6) = \{x_1, x_3\}$$
$$\mu_A(x_6) = \frac{4+3}{4+3+3+2+0+2} = 0.307$$
$$R(x_7) = \{x_2, x_4, x_6, x_7, x_8\},$$
$$A \cap R(x_7) = \{x_2\}$$

International Science Index, Mathematical and Computational Sciences Vol:12, No:3, 2018 waset.org/Publication/10008697

$$\mu_A(x_7) = \frac{4}{4+4+3+0+2} = 0.307$$

$$R(x_8) = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$$

$$A \cap R(x_8) = \{x_1, x_2, x_3\}$$

$$\mu_A(x_8) = \frac{4+4+4}{4+4+4+2+3+1+2+0} = 0.6$$

*Case II.* for $\lambda \leq 3$

$$R(x_1) = \{x_1, x_2, x_3\},$$

$$A \cap R(x_1) = \{x_1, x_2, x_3\}$$

$$\mu_A(x_1) = \frac{0+2+2}{0+2+2} = 1.0$$

$$R(x_2) = \{x_1, x_2, x_3\},$$

$$A \cap R(x_2) = \{x_1, x_2, x_3\}$$

$$\mu_A(x_2) = \frac{2+0+2}{2+0+2} = 1.0$$

$$R(x_3) = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

$$A \cap R(x_3) = \{x_1, x_2, x_3\}$$

$$\mu_A(x_3) = \frac{2+2+0}{2+2+0+3+2+3} = 0.33$$

$$R(x_4) = \{x_3, x_4, x_5, x_6, x_8\},$$

$$A \cap R(x_4) = \{x_3\}$$

$$\mu_A(x_4) = \frac{3}{3+0+1+3+2} = 0.33$$

$$R(x_5) = \{x_3, x_4, x_5, x_6, x_8\},$$

$$A \cap R(x_5) = \{x_3\}$$

$$\mu_A(x_5) = \frac{2}{2+1+0+2+3} = 0.25$$

$$R(x_6) = \{x_3, x_4, x_5, x_6, x_7, x_8\},$$

$$A \cap R(x_6) = \{x_3\}$$

$$\mu_A(x_6) = \frac{3}{3+3+2+0+3+1} = 0.25$$

$$R(x_7) = \{x_6, x_7, x_8\},$$

$$A \cap R(x_7) = \phi$$

$$\mu_A(x_7) = \frac{|\phi|}{3+0+2} = 0$$

$$R(x_8) = \{x_4, x_5, x_6, x_7, x_8\}$$

$$A \cap R(x_8) = \phi$$

$$\mu_A(x_8) = \frac{|\phi|}{2+3+1+2+0} = 0$$

*Case III.* for $\lambda \leq 2$

$$R(x_1) = \{x_1, x_2, x_3\},$$

$$A \cap R(x_1) = \{x_1, x_2, x_3\}$$

$$\mu_A(x_1) = \frac{0+2+2}{0+2+2} = 1$$

$$R(x_2) = \{x_1, x_2, x_3\},$$

$$A \cap R(x_2) = \{x_1, x_2, x_3\}$$

$$\mu_A(x_2) = \frac{2+0+2}{2+0+2} = 1$$

$$R(x_3) = \{x_1, x_2, x_3, x_5\},$$

$$A \cap R(x_3) = \{x_1, x_2, x_3\}$$

$$\mu_A(x_3) = \frac{2+2+0}{2+2+0+2} = 0.67$$

$$R(x_4) = \{x_4, x_5, x_8\},$$

$$A \cap R(x_4) = \phi$$

$$\mu_A(x_4) = \frac{|\phi|}{0+1+2} = 0$$

$$R(x_5) = \{x_3, x_4, x_5, x_6\},$$

$$A \cap R(x_5) = \{x_3\}$$

$$\mu_A(x_5) = \frac{2}{2+1+0+2} = 0.4$$

$$R(x_6) = \{x_5, x_6, x_8\},$$

$$A \cap R(x_6) = \phi$$

$$\mu_A(x_6) = \frac{|\phi|}{2+0+1} = 0$$

$$R(x_7) = \{x_7, x_8\},$$

$$A \cap R(x_7) = \phi$$

$$\mu_A(x_6) = \frac{|\phi|}{0+2} = 0$$

$$R(x_8) = \{x_4, x_6, x_7, x_8\}$$

$$A \cap R(x_8) = \phi$$

$$\mu_A(x_8) = \frac{|\phi|}{2+1+2+0} = 0$$

A comparison between the suggested method and previous method is shown in Table III.

TABLE III
COMPARISON BETWEEN SUGGESTED AND PREVIOUS METHODS

| | $x$ | The suggested method $\mu_A(x)$ | Pawlak $\mu_A(x)$ |
|---|---|---|---|
| Case I $\lambda \leq 4$ | $x_1$ | 0.250 | 0.488 |
| | $x_2$ | 0.250 | 0.428 |
| | $x_3$ | 0.250 | 0.430 |
| | $x_4$ | 0.520 | 0.375 |
| | $x_5$ | 0.625 | 0.428 |
| | $x_6$ | 0.307 | 0.330 |
| | $x_7$ | 0.307 | 0.800 |
| | $x_8$ | 0.600 | 0.375 |
| Case II $\lambda \leq 3$ | $x_3$ | 0.330 | 0.500 |
| | $x_4$ | 0.330 | 0.600 |
| | $x_5$ | 0.250 | 0.400 |
| | $x_6$ | 0.250 | 0.500 |
| Case III $\lambda \leq 2$ | $x_3$ | 0.670 | 0.750 |
| | $x_5$ | 0.400 | 0.500 |

## IV. Conclusion

The concept developed in this research presents a new method of finding the function of membership using the sum of degrees of symmetry and this leads to more accurate decisions in evaluating the approximations and decision making.

## References

[1] A. Elshenawy, M. Kamel Elsayed and E. Esodany, Weighted membership based on matrix relation, *Mathematical Methods in the Applied Sciences*, 2017.
[2] A. Oellicel, E. Leandor, G. Lambert, L. Eduardo and P. Rodrigues, "Intelligent Data Mining for Turbo-Generator Predictive Maintenance: *An Approach in Real-world*".
[3] E. Lashin, AM. Kozae, A. A. Abo Khadra, and T. Medhat, Rough set theory for topological spaces. *Int. J. Approx. Resoning*, 40, 35-43, 2005.
[4] K. Atanassov, Intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, 20, 87-96, 1986.
[5] L. A. Zadeh, Fuzzy Sets, *Inform. And Control*, 8, 338-353, 1965.
[6] R. Intan and M. Mukaidono, Generalization of rough membership function based on *α*-converings of the universe, *N. R. Pal and M. Sugeno* (Eds.), 129-136, 2002.
[7] R. Slowinski, D. Vanderpooten, A generalized definition of rough approximations based similarity, IEEE Trans. *Data Knowledge Eng*, 2, 331-336, 2002.
[8] T. Herawan and W. Meseri, Rough set membership function-based for clustering web transactions, *International Journal of Multimedia and Ubiquitous engineering*, 8, 105-118, 2013.
[9] Z. Pawlak, Rough Sets, *International Journal of Computer and Information Science*, 11 (5), 341-356, 1982.
[10] Z. Pawlak, Rough sets, theoretical aspects of reasoning about data: *Kluwer Academic Publishers*, 248, 1992.
[11] Z. Pawlak and A. Skowran, Rough membership functions: A tool for reasoning with uncertainty, *algebraic method in logic and in computer science*, 28, 135-149, 1993.
[12] Z. Huang, "A fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining".

International Science Index, Mathematical and Computational Sciences Vol:12, No:3, 2018 waset.org/Publication/10008697