WAS Science Nature
WASSN

ISSN: 2766 - 7715
WAS PRESS

Type of the Paper (Article, Review, Tutorial, Survey, Communication, etc.)

Reducing Fuzzy Data Set Attributes in Industrial Internet of Things (IIoT) Using Rough Set Theory

Mohamed A. Elashiri [1]*, Ahmed T. Shawky[2], and Abdelwahab S. Hassanc[3]

[1] Mohamed A. Elashiri is with Computer Science Dep., Faculty of Computers and Information System Beni-Suef University, Egypt
[2] Ahmed T. Shawky is with Management Information System Dep., El Madina Hight Institute of Administration and Technology, Egypt
[3] Abdelwahab said Hassan is with Math. And computer science department, Faculty of science, Suez Canal University, Egypt
*

Received January 14th, 2018; Accepted: March 30th, 2018; Published: April, 5th, 2018

Abstract: Due to the problem of attribute redundancy of fuzzy data set in the Industrial Internet of Things (IIoT) to simplify the induced decision rules without reducing the classification accuracy. In rough set theory, a reduct is generally defined as a minimal subset of attributes that can classify the same domain of objects as unambiguously as the original set of attributes. Based on fuzzy rough technique and using an efficient criterion in selection of fuzzy expanded attributes is important for reduction fuzzy data sets. In this paper proposes a new criterion, to reduce fuzzy attributes and keep of some attributes which selected by using accuracy measure of fuzzy expanded attributes with respect to fuzzy decision attributes.

Keywords: WASSN, Science, Advanced

1. Introduction

Digital transformation of smart manufacturing means production transforming using technologies like robotics, Internet of Things (IoT), Intelligent systems, and real-time analytics. Smart manufacturing is "fully-integrated, collaborative manufacturing systems that respond in real time to meet changing demands and conditions in the factory, in the supply network, and in customer needs" [1]Industrial internet of things (IIoT) uses sensors to collect data that would help gain efficiencies, speed up processes and reduce overall costs for a product or service. The hardware part contains some of the sensors, actuators and microcontrollers, while the functions of these sensors are perpetually recording and gathering data such as the natural biometrics measures for example temperature, ground wet, soil moisture sensor for detecting the temperature degree, amount of water in the soil and humidity. According to the data collected from this equipment the software part will apply a fuzzy computational algorithm to feature selection and reduct the fuzzy data set. Both fuzzy sets and rough sets theories are generalizations of classical set theory for modelling vagueness and uncertainty. Rough sets are the results of approximating crisp sets using equivalence classes, in other words, in traditional rough set approach, the values of attributes are assumed to be nominal data, i.e. symbols. Fuzzy set theory deal with the ill definition of the boundary of a class through a continuous generalization of set characteristic functions. In many applications, however, the attribute values can be linguistic terms (i.e. fuzzy sets), for example, the attribute ''height'' may be given the values of ''High'', ''Mid'', and ''Low'', then traditional rough set approach would treat these values as symbols, thereby some important information included in these values such as the partial ordering and membership degrees is ignored, which means that the traditional rough sets approach cannot effectively deal with fuzzy initial data e.g. linguistic terms. [2];[3];[4];[5];[6];[7];[13];[15];[17];[28],[29]. In this paper, we extend the concept of DTI dealing with meaningful fuzzy labels in order to express human knowledge for reducing fuzzy data set with using fuzzy rough technique.

The paper is organized as follows: Section 2 displays New Criterion for Selection of Expanded Attribute, in section 3, illustrates Differences between accuracy measure and dependency degree. The Procedure of Reducing Fuzzy Data Set using concept of Accuracy Measure is discussed in section 4. The experimental presents in Section 5, section 6 concludes the paper.

2. New Criterion for Selection of Expanded Attribute

Pawlak proposed two numerical measures for evaluating uncertainty of a set are accuracy and roughness [8] [14] [20] [21] [22] [24] [25], definition of accuracy measures was given, and suggested form in fuzzy rough set which is as follows: consider a non-leaf node S consisting of n attributes S={ F1 ,...,Fn} to be selected. For each k (1 ≤k≤ n), the attribute Fk takes mk values of fuzzy subsets, Fk = {F1k,......, Fmkk} and the fuzzy classification is FC ={ FC1,FC2,……,FCmc}, [9];[10];[11]; [12];[16];[23];[26];[27] as shown in Table 1.

Table 1. Structure of fuzzy information System FIS

U	F^1			F^k			F^{Class}		
	$F^1{}_1$	.	$F^1{}_{m1}$	$F^k{}_1$	.	$F^k{}_{mk}$	$F^C{}_1$	.	$F^C{}_{mc}$
X_1	$\mu_{F_1^1}(x^1)$	.	$\mu_{F_{m1}^1}(x^1)$	$\mu_{F_1^k}(x^1)$	.	$\mu_{F_{m1}^k}(x^1)$	$\mu_{F_1^C}(x^1)$	.	$\mu_{F_{m_c}^C}(x^1)$
:	:	:	:	:	:	:	:	:	:
X_N	$\mu_{F_{m1}^1}(x^N)$	.	$\mu_{F_{m1}^1}(x^N)$	$\mu_{F_{m1}^k}(x^N)$	.	$\mu_{F_{m_k}^k}(x^N)$	$\mu_{F_1^C}(x^N)$	.	$\mu_{F_{m_c}^C}(x^N)$

Definition 1 The accuracy measure is equal to completeness of knowledge about the given object set x is defined by the ratio of cardinality of lower and upper approximation set of x as [18]; [27]:

$$\alpha_R(x)=\frac{\left|\underline{R}(x)\right|}{\left|\overline{R}(x)\right|} \quad (1)$$

Definition 2 Let U be a given universe. R is a fuzzy equivalence relation over U. The fuzzy equivalence class [x]R is defined by:

$$\mu_{[x]_R} = \mu_R(x,y) \quad (2)$$

Definition 3 The one of important issue in data analysis is discovering dependencies between attributes; set of attributes D depends totally on a set of attributes C, if all values of attributes from D are uniquely determined by values of attributes from C, formally dependency can be defined as:

$$\gamma(C,D)=\frac{|POS_C(D)|}{|U|} \quad (3)$$

Definition 4 Let U be a given universe, for arbitrary class l from fuzzy classification FC={ FC1,FC2,……,FCmc} and fuzzy partitions Fjk∣ 1≤j≤mk, 1≤k≤n}, fuzzy rough set is a tuple ($\underline{\mu}_l$,

$\mu_{\bar{l}}$) where $\mu_{\underline{l}}$ is lower approximation membership degree and $\mu_{\bar{l}}$ is upper approximation membership degree of class l through Fjk are define by [Dubois , 92]:

$$\mu_{\underline{l}}\left(F_{jk}\right)=\inf_{\forall i\in u}\ \max\left\{1-\mu_{F_{jk}}(x_j^i),\mu_l(y^i)\right\} \qquad (4)$$

$$\mu_{\bar{l}}\left(F_{jk}\right)=\sup_{\forall i\in u}\ \min\left\{\mu_{F_{jk}}(x_j^i),\mu_l(y^i)\right\} \qquad (5)$$

Definition 5 (Proposed) The accuracy measure is equal to completeness of knowledge about the fuzzy partition Fjk is defined by the ratio of lower and upper approximation membership degree of class l, suggested forms of the fuzzy rough accuracy measure in fuzzy data set and fuzzy rough set is:

$$\alpha_l\left(F_{jk}\right)=\frac{\mu_{\underline{l}}(F_{jk})}{\mu_{\bar{l}}(F_{jk})} \qquad (6)$$

Property 1: (Maximum) Let SF = (U, F) be an fuzzy information system, R⊆F, and X ⊆ U, the maximum accuracy of Fkj with respect to R is 1.

Property 2: (Minimum) Let S = (U, A) be an information system, R⊆F, and X⊆ U, the minimum accuracy of Fkj with respect to R is zero.

Property 3: If $\mu_{\underline{l}}\left(F_j^k\right)=0$, then $\alpha_l\left(F_j^k\right)$ accuracy measure is zero, regardless of the size of the fuzzy upper approximation.

Definition 6: Fuzzy rough accuracy degree the proposed accuracy measure degree of fuzzy attribute Fk of class l is:

$$\alpha_l(F^k)=\sum_{j=1}^{m_k}\left(\frac{M\left(F_j^k\right)}{\sum_{j=1}^{m_k}M\left(F_j^k\right)}\right)\alpha_l\left(F_j^k\right) \qquad (7)$$

Where α (Fki) is fuzzy Accuracy measure of sub attribute Fki and the term M(Fkj) /∑mkj=1 M(Fkj), is the weights which represent the relative size of subset Fkj in F. The attribute with maximum value of accuracy measure αl(Fk) will be selected as a root of fuzzy decision tree i.e. Root=Maxk(αl (Fk)) $\in$ Sroot , 1≤k≤n

3. Differences between accuracy measure and dependency degree

The main differences between accuracy measure and dependency degree in both views of crisp and fuzzy as the following:

1. In Equation (1) the main difference between dependency degree and accuracy measure in both formulas is the demonstrator in the first one is the cardinality of U; set of objects.
2. The cardinality of upper approximation of the set. In addition:
    a. The cardinality of upper approximation of the set Less than or equal of is the cardinality of U
    b. The cardinality of U is constant for all attributes within data set, so if the numerator is constant also for two attributes then directly the dependency degree will be equal.
    c. The accuracy measure gives the chance for all sub attributes to describe itself.

Property 1: Set of numerators of αP(x) $\subseteq$ Set of numerators of γD(x). By let $x \in$ numerator of αP (x) $\Leftrightarrow$ $x \in$ SP(x) $\Leftrightarrow$ $x \in$ PoSP(x) $\Leftrightarrow$ $x \in$ γP(x).

Property 2: αP (x) = 1 $\Leftrightarrow$ γP(x) = 1.

Property 3: αP (x) = 0 $\Leftrightarrow$ γP(x) = 0.

4. The Procedure of Reducing Fuzzy Data Set using concept of Accuracy Measure

1- Initial fuzzy data set: first read two files one of them to description the structure of all attribute's conditionals and decisions, the second of them for the data. Next, transform the fuzzy data into decision table format.

2- Fuzzy attribute reduction: compute the attribute reduct of initial fuzzy dataset using accuracy measure degree in fuzzy rough form.

3- Reducing fuzzy attribute-values: Remove the superfluous and irrelevant attribute values from the dataset by using accuracy measure degree.

Step 1. Pre-paring the FIS
Step 2. Fuzzy reduct attribute = { }, αbest =0, αprev =0, Tempatt ={}
Step 3. Do
Step 4. Tempattributes = Fuzzyreductattribute
Step 5. αprev = αbest
Step 6. For τ attributes in (Conditional attributes- Fuzzyreductattribute)
Step 7. Calculate the Accuracy measure $\alpha(\tau \cup$Fuzzyreductattribute) using equation (7).
Step 8. Calculate the Accuracy measure α(Tempattributes) using equation (7).
Step 9. If $\alpha(\tau \cup$Fuzzyreductattribute) = 1 then
Step 10. Fuzzyreductattribute =$\tau \cup$Fuzzyreductattribute

Step 11. Return Fuzzyreductattribute
Step 12. End if
Step 13. If(α(τ∪Fuzzyreductattribute)< α(Tempattributes)) then
Step 14. Go to step 6
Step 15. Else
Step 16. Tempattributes=τ∪ Fuzzyreductattribute
Step 17. αprev = α(Tempattributes)
Step 18. Fuzzyreductattribute = Tempattributes
Step 19. Until (αbest = αprev)
Step 20. Return Fuzzyreductattribute
Step 21. End

Definition 7 : Gamma is the Degree of cutting reduct, Gamma factor degree of accuracy measure of attributes, can cut the process of growing the reduct as if the accuracy measure less than Gamma then continues otherwise the growing will stopped.

4. The Algorithm of the Reduct Fuzzy Decision Tree.

The algorithm aims to reduce the conditional fuzzy attributes from fuzzy information system, based on maximum Accuracy function of each conditional attribute in dataset, see equation (6). Figure (1) shows the flow chart of the proposed algorithm.

4.1. The algorithm for Reducing Fuzzy information using accuracy measure

This section introduces the algorithm for reducing fuzzy data set based on accuracy measure; the process consists of the following steps:

Consider a non-leaf node S consisting of n attributes F1,..., Fn, to be selected for each k (1 ≤ k ≤ n), the attribute Fk takes mk values of fuzzy subsets, F1k,...... Fmk, as :.

Input: Fuzzy Information system

(C:condiaiton attributes ,D decision attributes)

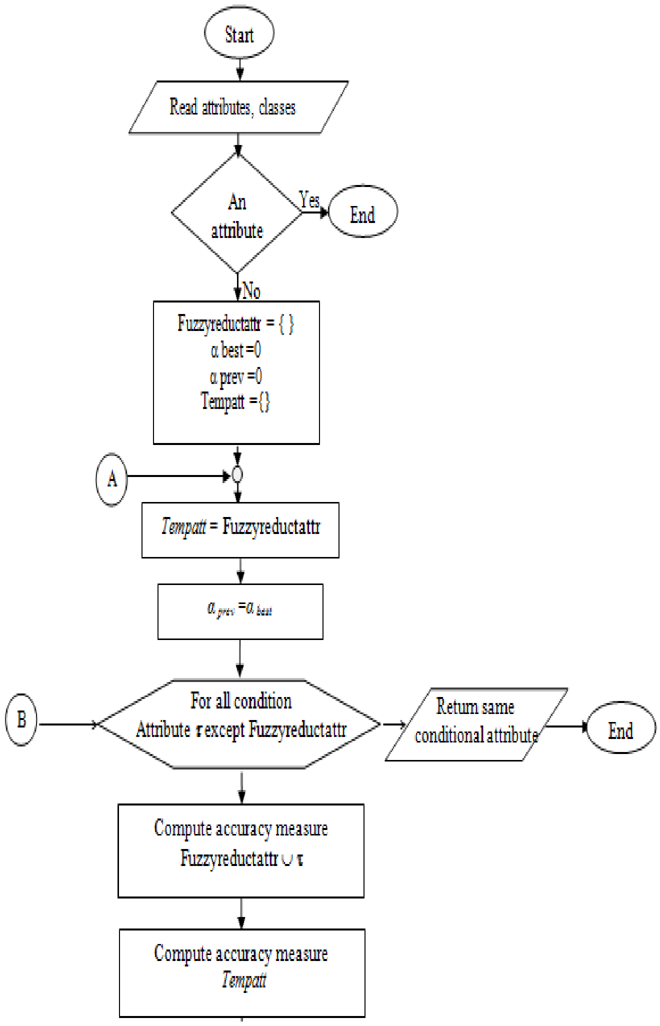
Output: Fuzzy Information system after reduct.

Pre-paring the FIS:

Fuzzy reduct attribute = { }, αbest =0, αprev =0, Tempatt ={}

The effectiveness of our technique is demonstrated through numerical experiments in the environment of C#. Our experiments select 5 datasets are from UCI [19];[21] iris is the data set which applied out proposed on it contains 150 records and 4 columns with 3 classes

The black column displays numbers of attributes after reduct depend on our proposed accuracy measure in range of gamma factor from 0.0 to 1.0 and the gray column according to dependency degree both are display numbers of attributes after reduct in range of gamma factor from 0.0 to 1.0.


Start
Read attributes, classes
An attribute
Yes
End
No
Fuzzyreductattr = { }
α best =0
α prev =0
Tempatt ={}
A
Tempatt = Fuzzyreductattr
α prev =α best
B
For all condition
Attribute τ except Fuzzyreductattr
Return same conditional attribute
End
Compute accuracy measure
Fuzzyreductattr ∪ τ
Compute accuracy measure
Tempatt

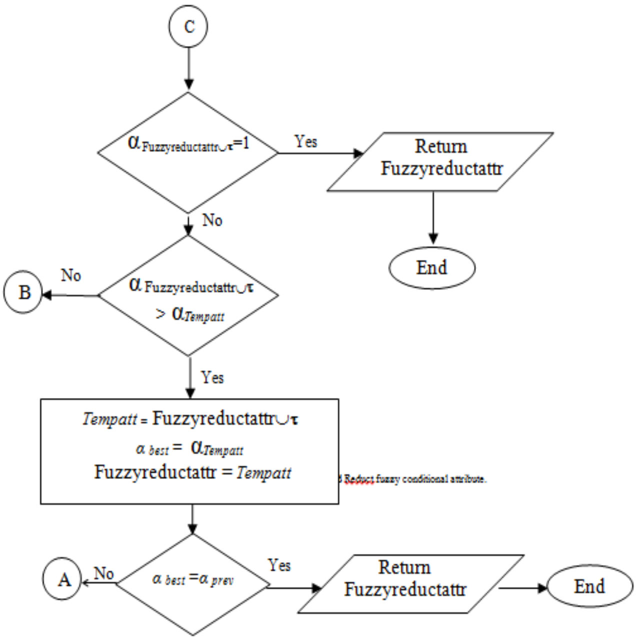


Figure 1. Flow chart of the algorithm of the proposed Reduct fuzzy conditional attribute.

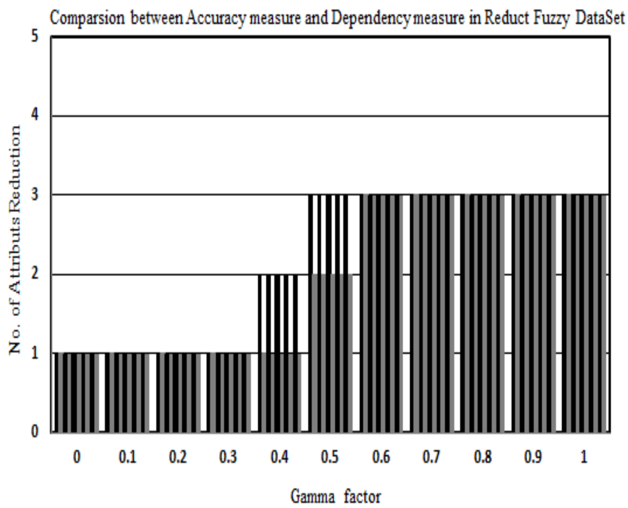


Figure 2. Comparison between Accuracy measure and Dependency measure in Reduct Fuzzy DataSet (numbers of attributes after reduct process)

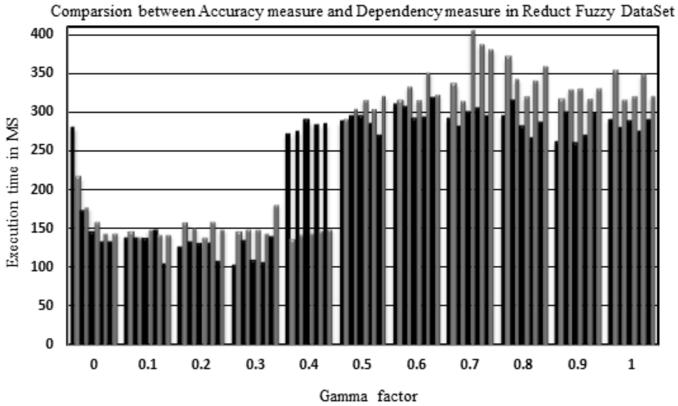


Figure 3. Comparison between Accuracy measure and Dependency measure in Reduct Fuzzy DataSet (Time in Ms of reduct process)

Author Contributions: The authors contributed equally to this research.

Funding: The authors receive no fund for this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. D. Zhang, C. Chuen and G. Zhou, "Enabling Industrial Internet of Things (IIoT) towards an emerging smart energy system, Global Energy Interconnection, vol. 1, issue 1, p.p. 39-47, 2018. https://doi.org/10.14171/j.2096-5117.gei.2018.01.005.
2. W.W. Zhi, J.S. Mi and W.X. Zhang "Generalized fuzzy Rough Sets," information Sciences, vol.151, p.p. 263–282, 2003.
3. X.Z. Wang, E.C.C. Tsangb, S. Zhao, D. Chen and D.S. Yeung, "Learning fuzzy rules from fuzzy samples based on rough set technique," information Sciences, vol. 177, p.p. 4493–4514, 2007.
4. T. Wang, Z. Li, Y. Yan, and H. Chen," A Survey of Fuzzy Decision Tree Classifier Methodology," Fuzzy information and Engineering (ICFIE), ASC 40, pp. 959–968, 2007.
5. Y. Ning, L. Tianruii and S. Jing "Construction of Decision Trees based Entropy and Rough Sets under Tolerance Relation," International Conference on Intelligent Systems and Knowledge Engineering (ISKE Oct. 2007). URL "http://www.atlantis-press.com/php/download_paper.php?id=1485"
6. Z. Huang, T. D. Gedeon and M. Nikravesh, "Pattern Trees Induction: A New Machine Learning Method," IEEE Transactions on Fuzzy Systems, vol. 16 no. 4, August 2008.
7. X.Z. Wang, J.H. Zhai and S.Xia Lu, "Induction of multiple fuzzy decision trees based on Rough Set technique," information Sciences, vol. 178 p.p. 3188–3202, 2008.
8. S.Ruying, "Data Mining Based on Fuzzy Rough Set Theory and Its Application in the Glass Identification," Modern Applied Science, Vol 3, No 8, 2009
9. Y. Yuan and M. J. Shaw, "Induction of fuzzy decision trees," Fuzzy Sets and systems, Vol. 69, p.p.125-139, 1995.
10. B. Rajen and M.Gopal "FRCT: fuzzy-rough classification trees," Pattern Analysis & Applications, vol. 11, Issue 1, p.p.73-88, 2008
11. X.Z. Wang, J.H. Yan and R. Wang "A sample selection algorithm in fuzzy decision tree induction and its theoretical analyses," IEEE International Conference on systems, Man and Cybernetics, p.p. 3621 – 3626, 7-10 Oct. 2007
12. J. Zhai , " Fuzzy decision tree based on fuzzy-rough technique," Soft Computing - A Fusion of Foundations, Methodologies and Applications, Springer-Verlag 2010 DOI:10.1007/s00500-010-0584-0
13. C.Z. Janikow "Fuzzy Decision Forest," Fuzzy information Processing Society, 2003. NAFIPS 2003. 22nd International Conference of the North American, p.p. 480 – 483. , 4-26 July 2003
14. H. H. Qing, Y. G. MING and R.Y. DA" Construct rough decision forests based on sequentially data reduction, proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, 13-16 Au gust 2006.

15. L. Xue, H.Z. Xiao and D.Z. Dong "Four Matching Operators of Fuzzy Decision Tree Induction," Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007) vol.4, pages 674-678, 2007.
16. X.Z. Wang and X.H. Gao "A Research on the Relation Between Training Ambiguity and Generalization Capability," International Conference on Machine Learning and Cybernetics, p.p. 2008-2013, Aug. 2006.
17. Elashiry M. "Fuzzy decision trees approach for data mining," Master thesis, Egypt, Cairo University, Institute of Statistical Studies and Researches, 2006 unpublished.
18. J. Liang, J. Wang and Y.Qian "A new measure of uncertainty based on knowledge granulation for rough sets," information Sciences, vol. 179, p.p. 458-470, 2009.
19. Blake CL, Merz CJ, UCI Repository of machine learning databases, 1996. http://www.ics.uci.edu/*mlearn/MLRepository.html
20. Z. Pawlak, A. Skowron, "Rudiments of rough sets," Information Sciences, vol. 177, p.p. 3-27, 2007.
21. X.Z. Wang, J.H. Zhai and S.F. Zhang, "Fuzzy decision tree based on the important degree of fuzzy attribute" Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, p.p. 511-516, 12-15 July 2008.
22. Y.L. Chen, H. W. Hu and K. Tang, "Construct a decision tree from data with hierarchical class labels" Expert Systems with Applications, vol. 36, p.p. 4838-4847, 2009
23. Elashiry M. A., H. A. Hefny, and A. H. Abdel Wahab, "Fuzzy Rough Decision Tree based on Rough Accuracy Measure and Fuzzy Rough Set", Proceedings of the 1st International Conference on Advanced Computing And Communication (ICACC 2010), Amal Jyothi College of Engineering , Kanjirapally, Kerala, India, 3-4 May 2010, pp.134-138.
24. H. A. Hefny, Ahmed S. Ghiduk , A. H. Elwahab and M. A. Elashiri "Effective Method for Extracting Rules from Fuzzy Decision Trees based on Ambiguity and Classifiability", Journal: Universal Journal of Computer Science and Engineering Technology ISSN/EISSN: 22192158 Year: 2010 Volume: 1 Issue: 1 Pages: 55-63, ISSN: 22192158 , available on" http://www.unicse.org/october2010/Effective_Method_for_Extracting_Rules_from_Fuzzy_Decision_Trees_based_on_Ambiguity_and_Classifiability.pdf".
25. M. A. Elashiri, H. A. Hefny, and A. H. Elwahab, "Construct Fuzzy Decision Trees Based on Roughness Measures". Advances in Communication, Network, and Computing, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering Volume 108, 2012, pp 199-207.
26. M. A. Elashiri, H. A. Hefny, and A. H. Elwahab, "based on fuzzy Introduction of fuzzy decision trees rough set techniques," In Proc. IEEE International Conference on Computer Engineering and Systems, 2011, pp. 134-139.
27. M. A. Elashiri, H. A. Hefny, and A. H. Elwahab. "Reduction Fuzzy Data Set based on Rough Accuracy Measure." In International Conference on Advances in Computer Science, AETACS. Elsevier. 2013.
28. Y.Y. Yao "A comparative study of Fuzzy Sets and Rough Sets," information Sciences, vol. 109,no. 1-4, pp. 227-242, 1998.
29. M. Masurah, A. Selamat, I. Subroto and O. Krejcar "Improving the classification performance on imbalanced data sets via new hybrid parameterisation model," Journal of King Saud University - Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2019.04.009, 2019.