International Journal of Natural and Engineering Sciences 5 (2): 69-72, 2011
ISSN: 1307-1149, E-ISSN: 2146-0086, www.nobel.gen.tr

Effects of Fuzzification on Fuzzy Rough Attribute Reduction (FRA) Algorithm

Bekir AGIRGUN[1] Cevriye GENCER[2] Emel KIZILKAYA AYDOGAN[3]

[1] Military Academy, Institute of Defense Science

[2] Department of Industrial Engineering, Faculty of Engineering, Gazi University, Ankara, TURKEY

[3] Department of Industrial Engineering, Faculty of Engineering , Erciyes University, Kayseri, TURKEY

*Corresponding Author
e-mail: bekiragi@gmail.com

Received: April 10, 2011
Accepted: April 19, 2011

Abstract

Over the last decades, numerous researches have been made on hybridization of fuzzy and rough sets. In this paper, it is purposed to evaluate the effects of fuzzification on the algorithm FRA which can find reducts in a fuzzy database using fuzzy rough set approach based on discernibility matrix. The FRA algorithm deals with more general type of fuzzy sets.

Keywords: Fuzzy Rough Sets, Discernibility Matrix, Attribute Reduction, more general type of Fuzzy Sets, effects of fuzzification.

INTRODUCTION

In recent years, the availability of large amount of data requires the extraction of useful information. Through technological developments, this kind of data can be collected from vast resources such as production control and business management.

Nowadays in the real world, one of the most important steps in extracting useful information from vast amount of data which is expressed in gigabytes is dimensional reduction of attributes. Through dimensional reduction of attributes, it is purposed to have some improvements in efficiency such as; decreasing measurement, storage and computation costs, increasing the performance of classification algorithms as well as better understanding the results of classification algorithms [1].

Feature selection introduces an active research area for researchers in pattern recognition, statistics and data mining communities.

Modeling of uncertainty provides a fruitful area for researchers about knowledge representation. Plenty of approaches like Zadeh's fuzzy set theory [2] and Pawlak's rough set theory [3] point out the uncertainty problem. Fuzzy rough sets are the generalization of classical rough set theory for modeling uncertainty. These two theories are different but complement and related to each other.

The Pawlak's rough set concept is a new mathematical approach to imprecision, vagueness and uncertainty. In rough set approach, an arbitrary subset of the universe of discourse can be approximated by two subsets by means of the equivalence classes; namely, lower and upper approximations. Through the lower and upper approximations, one can not only extract the decision rules that are hidden in the database but also select the minimal subset of data that is the most informative.

Dubois and Prade [4,5], are one of the first who investigated the problem of fuzzification of a rough set. In their research, they constructed the lower and upper approximation by means of operators t-norm min and t-conorm max.

In the fuzzy rough set theory, fewer efforts have been put on the attribute reduction in fuzzy databases [6]. In this context, one of the important studies on fuzzy rough set approach is the fuzzy-rough QuickReduct algorithm that has been proposed by Jensen and Chen [7-9]. Another important study about attribute reduction on fuzzy rough set concept has been made by Salido and Murakami [10]. In their paper, they defined the more general type of fuzzy sets and proposed a strong infrastructure about attribute reduction applications based on fuzzy rough sets that can be used by constructive approaches. In their approach, they used t-similarity relation which is reflexive, symmetric and have a certain transitivity property. Tsang *et al*. [11] proposed an algorithm based on discernibility matrix that can find all the reductions on a fuzzy database. However, in their algorithm, t-norm min operator was used for the purpose of constructing and aggregating fuzzy similarities. Aydogan *et al*. [12] purposed an algorithm that combines all the features of these approaches. Their algorithm first fuzzifies the continuous and categorical values of a database, constructs the fuzzy similarity matrices according to a certain type of t-norm, aggregates the similarity matrices and finally constructs the discernibility matrix that can find all the reductions of that fuzzy database.

In this paper, algorithm in [12] is compared with the other well-known attribute reduction algorithms and it is purposed to find a value range of fuzzification parameter σ that affects the classification accuracies and the minimal reducts.

The rest of the paper is organized as follows: In the Material and Methods Section; the basic and the general type of fuzzy sets, the transformations of notations that are used in the algorithm, the FRA algorithm and the fuzzification of datasets are defined. In Results and Discussion Section; the efficiency of the FRA algorithm and the effects of fuzzification parameter on the classification accuracies of datasets are discussed. Conclusions are given in final section.

MATERIALS AND METHODS

The literature of fuzzy-rough sets usually deals with simpler type of fuzzy sets. However, some mathematical transformations are necessary in order to perform calculations on general type of fuzzy sets. In this section, information on notations and the algorithm will be given.

Basic type of fuzzy sets[10]

A data set with N objects which is described by N_F characteristics or features that are fuzzy. In this data set, every sample x_i ($i \in [1, N]$), have been measured as partial membership degrees, μ_i^j , $j = 1,.., N_F$, $x_i = [\mu_i^1, \mu_i^2, ..., \mu_i^{N_F}]$ which are graded in the interval [0,1].

General type of fuzzy sets[10]

A data set of N samples in which sample x_i ($i \in [1, N]$) is described by N_F features, while the value of each feature j is expressed by a set of N_{LL} grades of membership, μ_i^{jk} $(k = 1,..., N_c L)$, to N_{LL} linguistic labels. Thus sample x_i can be characterized by the following values;

$$x_i = [(\mu_i^{11}, ..., \mu_i^{1k}, ..., \mu_i^{1N_{LL}}) \; (\mu_i^{21}, ..., \mu_i^{2k}, ..., \mu_i^{2N_{LL}}), \\ ...., (\mu_i^{j1}, ..., \mu_i^{jk}, ..., \mu_i^{jN_{LL}}), ..., (\mu_i^{N_F 1}, ..., \mu_i^{N_F k}, ..., \mu_i^{N_F N_{LL}})]$$

where $i = 1,..., N,\; j = 1,..., N_F,\; k = 1,..., N_{LL}$,

and $\mu_j^i \in [0,1]$. In this paper we assume the same number of linguistic labels, N_{LL} , for all features.

Fuzzification of Data Sets

In order to fuzzify the datasets, a simple method that only uses the information in the dataset itself was chosen. For numerical or continuous values, the value of every measured attribute F_i , is represented as a normalized attribute value F_i^* where $F_i^* = (F_i - \mu_i) / F_i$. Here μ_i is the mean of F_i in the whole attribute set.

After the normalization process, every attribute is fuzzified and transformed into a set of membership degrees represented by fuzzy partition in Figure 1. The parameters defining the membership functions take a value which depends on the variance of the attribute set. The parameter 0,9 σ was randomly chosen before making any tuning.

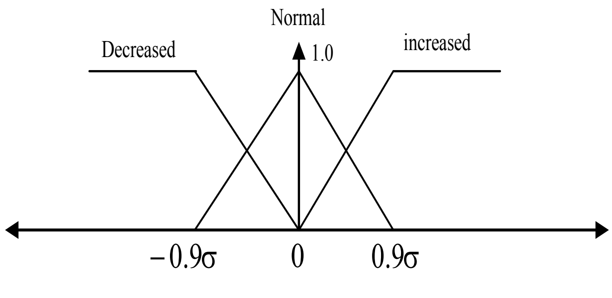

Figure 1. Fuzzy partition used to express every attribute.

The Fuzzy Rough Attribute Reduction (FRA) Algorithm

The attribute reduction with fuzzy rough sets based on the constructive approach has been introduced in [12]. Here, some adaptations on constructing the fuzzy similarity matrices and the aggregating operator are given for the general type of fuzzy sets. Then, the indiscernibility matrix is formed. Finally all the reducts from the indiscernibility matrix are found.

In order to find the fuzzy similarity matrices $Sim(R_i)$, the Lukasiewicz operator;

$$T(x, y) = Max(x + y - 1, 0)$$

is turned into $S_{ij}^k(x_i, y_j) = \inf_{l \in J_{ij}^k} (1 - |\mu_i^{kl} - \mu_j^{kl}|)$,

The De Morgan dual of generalized means,

$$M(x) = 1 - (((1 - x_1)^p + (1 - x_2)^p + ... + (1 - x_n)^p) / n)^{1/p}$$

is chosen as the aggregation operator for T -fuzzy similarity matrices. Depending on how parameter p is tuned, this operator behaves like a minimum or a mean average operator. It is better to choose $p = 2$ as the minimum operator as reported in [10].

Now we can define attributes reduction for fuzzy rough sets using these views as follows;

Let U is a finite universe of discourse and $\Re$ is a finite set of fuzzy t-similarity relations called conditional attributes set. D is an equivalence relation called decision attribute with symbolic values. $(U, \Re \cup D)$ is called t-fuzzy decision system.

Denote $Sim(\Re) = \cap\{R : R \in \Re\}$,

then $Sim(\Re)$ is also a fuzzy t-similarity relation. Suppose $[x]_D$ is the equivalence class with respect to D for $x \in U$, then the positive region D relative to $Sim(\Re)$ is defined as;

$$Pos_{Sim(\Re)} D = \cup_{x \in U} Sim(\Re)_{\vartheta}([x]_D).$$

We will say that $\Re$ is dispensable relative to D in $\Re$ if $Pos_{Sim(\Re)} D = Pos_{Sim(\Re - \{R\})} D$, otherwise we will say that $\Re$ is indispensable relative to D in $\Re$. The family $\Re$ is independent relative to D if each $R \in \Re$ is indispensable relative to D in $\Re$; otherwise $\Re$ is dependent relative to D . $P \subset \Re$ is a reduct relative to D if P is independent relative to D and $Pos_{Sim(\Re)} D = Pos_{Sim(P)} D$. P will be called relative reduct of $\Re$. The collection of all indispensable elements relative to D in $\Re$ is called the core of $\Re$ relative to D , denoted as $Core_D(\Re)$. Similar to the result in traditional rough sets, $Core_D(\Re) = \cap \text{Red}_D(\Re)$, $\text{Red}_D(\Re)$ is the collection of all the relative reducts of $\Re$. Thorough, it is known that $\{R_T x_\lambda : x \in U, \lambda \in (0,1]\}$ could be basic granular set to construct lower and upper approximations of fuzzy sets since every lower and upper approximation is just the union of the fuzzy sets with the form as $R_T x_\lambda$.

Thus, the structure of lower approximation of every $[x]_D$ is clear by $\underline{R}_{\vartheta}([x]_D) = \cup\{R_T(y_\lambda) : R_T(y_\lambda) \subseteq [x]_D\}$ For $y \notin [x]_D$, clearly $\underline{R}_{\vartheta}([x]_D)(y) = 0$ holds. For $y \in [x]_D$, the following theorem develops a sufficient and necessary condition for $\overline{R}_T(y_\lambda) \in \underline{R}_{\vartheta}([x]_D)$.

Theorem 1 [13].

Suppose $y \in [x]_D$, $\overline{R_T}(y_\lambda) \subseteq \underline{R}_\vartheta([x]_D)$ if and only if $\overline{R_T}(y_\lambda)(z) = 0$ for $z \notin [x]_D$.

Theorem 2 [13].

Suppose $P \subset \Re$, $Pos_{Sim(\Re)}D = Pos_{Sim(P)}D$ if and only if $Sim(P)_T(x_{\lambda(x)}) \subseteq [x]_D$ for every $x \in U$, here $\lambda(x) = \underline{Sim(\Re)}_\vartheta([x]_D)(x)$.

Theorem 3 [13].

Suppose $P \subset \Re$, then P contains a relative reduction of $\Re$ if and only if $\overline{Sim(P)_T}(x_{\lambda(x)})(z) = 0$ for every $x, z \in U$ and $x \notin [x]_D$. Here $\lambda(x) = \underline{Sim(\Re)}_\vartheta([x]_D)(x)$.

Theorem 4 [13].

Suppose $P \subset \Re$, then P contains a relative reduction of $\Re$ if and only if there exists $p \in P$ such that $T(P(x,z)\ \lambda(x)\ = 0$ for every $x, z \in U$ and $x \notin [x]_D$ With above discussion we can design an algorithm to compute all relative reductions. Suppose $U = \{x_1, x_2, ..., x_n\}$, $\Re = \{R_1, R_2, ..., R_m\}$. By $M_D(U, \Re)$, a nxn matrix, denoted (c_{ij}), called discernibility matrix of $(U, \Re \cup D)$ such that;

1. $c_{ij} = \{R \in \Re : T(R(x_i, x_j\lambda(x_i) = 0\}$ if $x_j \notin [x_i]_D$;

2. $c_{ij} \neq 0$, otherwise.

Here $M_D(U, \Re)$ may not be symmetric and $c_{ii} = \phi$.

Reduct

Let $M_D(U, \Re)$ denotes a nxn matrix (c_{ij}), called discernibility matrix of $(U, \Re \cup D)$. By M_H we denote the matrix from $M_D(U, \Re)$ eliminated all the elements whose intersection with set $\Re$ is non-empty. count(a) represents the times that attribute "a" appears in matrix $M_D(U, \Re)$ [14]. So the significance of an attribute "a" denoted by $SGF(a, \Re, P) = count(a)$. Generally, the bigger the $SGF(a, \Re, P)$, the more important the attribute in $M_D(U, \Re)$.

The FRA algorithm

Input: $(U, \Re \cup D)$ fuzzy decision system,
Output: P, reducts of conditional attributes.

Step 1. Compute $Sim(R_1), ..., Sim(R_n)$

Step 2. Compute
$Sim(\Re) = M(Sim(R_1), ..., Sim(R_n)),$

Step 3. Compute $M_D(U, \Re)$ such that
$c_{ij} = \{R : 1 - R(x_i, x_j) \geq \lambda_i\}$

Step 4. Delete the entries in $M_D(U, \Re)$ such that $c_{ij} = \emptyset$

Step 5. Sort rest of c_{ij}'s according to their cardinality. Cardinality (size of a set) shown as

L_i, $(i = 1, 2, ..., m)$,

Step 6. Let "a" be a conditional attribute and P is the reduct set. Compute the significance of attribute(s) at a certain subset cardinality $(SGF(a', \Re, P) \mid L_{i_0}) = Max_{a \in P-R}\{SGF(a, \Re, P \mid L_{i_0}\}$.

Step 7. $P \Leftarrow P \cup \{a'\}$ such that
$M_D(U, \Re) = \emptyset$

RESULTS AND DISCUSSION

In this section, the FRA algorithm was compared with the other well known attribute reduction algorithms such as classical rough set method and fuzzy-rough QuickReduct algorithm. Furthermore, the effects of fuzzification on fuzzy rough attribute reduction (FRA) algorithm were investigated. In order to measure the efficiency of reducts, Naïve Bayes [15] algorithm was used.

Experimental data sets were taken from UCI machine learning repository [16] and four different datasets were chosen. Table 1 shows data descriptions. Classification performance of selected attributes was evaluated by using Naïve Bayes algorithm. In order to calculate classification accuracies of datasets, Waikato Environment for Knowledge Analysis (WEKA) [17] program was used. Besides, in order to find the reducts of classical rough set approach, ROSETTA [18] program was used. Boolean Reasoning algorithm was used for discretization of data sets. Additionally, p=2 index for aggregating fuzzy similarity matrices was used. Table 2 shows the selected attribute set cardinality. Attribute set cardinality is closely related to the computational time of the classification algorithm. Minimal cardinalities point out the efficiency of the algorithm. Furthermore, Table 3 shows the Naïve Bayes classification algorithm accuracies of the reduced datasets. The percentage of accuracy shows the efficiency of the algorithm.

Effect of the fuzzification parameter σ

In this section, it was investigated how the classification accuracy has been affected by the fuzzification parameter σ. After tuning the σ parameter, it was observed that cardinality of the selected attributes and the classification accuracies were changed. As shown in Table 4, selecting σ in the range of [0.8-0.9] while using the corresponding datasets and Naïve Bayes classification algorithm provides smaller cardinalities and higher classification accuracy results.

Table 1. Data Descriptions

No	Dataset	Samples	Numerical Features	Categorical Features	Classes
1	Credit Approval	690	6	9	2
2	P.I.Diabetes	768	8	0	2
3	Ecoli	336	5	2	7
4	WPBC	198	33	0	2

Table 2. Attribute set cardinality after feature selection

Dataset	Selected attribute set cardinality Nr of attributes	Rough Reduct	FR Quick-Reduct	FRA Algorithm
Credit Approval	15	12	14	8
P.I.Diabetes	8	7	8	5
Ecoli	7	6	6	5
WPBC	33	1	6	2
Average	15.75	6.5	8.5	**5**

Table 3. Naïve Bayes Classification Algorithm Accuracies of Selected Attributes.

Dataset	Naive Bayes Classification Accuracy			
	All Attributes	Rough Reduct	FR Quick-Reduct	Proposed Algorithm
Credit Approval	77.5362	76.087	77.5362	84.7826
P.I.Diabetes	76.5625	76.5625	76.5625	77.2135
Ecoli	84.2262	84.8214	85.4167	86.0119
WPBC	63.4021	74.7423	76.2887	81.4433
Average	75.43175	78.0533	78.95103	**82.36283**

Table 4. The effect of fuzzification parameter to the classification accuracies of datasets.

σ **Value**	**Classification Accuracies (**Naïve Bayes)			
	Credit Approval	P.I.Diabetes	Ecoli	WPBC
0.6	83.1884 [7]	**77.2135 [6]**	82.5926 **[4]**	77.8351 [3]
0.8	**85.5072 [7]**	**77.2135 [6]**	80.7407 [5]	78.3505 [3]
0.85	84.3478 [10]	77.0833 [6]	**83.3333** [5]	76.2887 **[2]**
0.9	84.7826 [8]	**77.2135 [6]**	**83.3333** [5]	**81.4433** [4]
0.95	84.7826 [8]	**77.2135 [6]**	76.2963 [5]	77.3196 **[2]**
0.99	84.7826 [9]	77.0833 [6]	79.6296 [5]	76.2887 [3]

CONCLUSION

In this paper, the efficiency of the FRA algorithm and the effects of tuned fuzzification parameter on the classification accuracies were evaluated. It was shown that the FRA algorithm provides both smaller set of attributes and better classification results. It is simple and easy to understand compared with the other fuzzy rough attribute reduction algorithms. In the future, the FRA algorithm will be used for a real world application.

REFERENCES

[1] Liu H., Lei Y., Motoda H. 2003. Feature extraction, selection and construction. Handbook of Data Mining. Lawrence Erlbaum Associates, Inc. 410.

[2] Zadeh, L.A. 1965. Fuzzy sets. Inform. Control. 8:338–353.

[3] Pawlak, Z. 1982. Rough sets. International Journal of Computer and Information Sciences. 11:341–356.

[4] Dubois, D. and Prade, H. 1990. Rough fuzzy sets and fuzzy rough sets. International Journal of General Systems. 17: 191–209.

[5] Dubois, D. and Prade, H. 1992. Putting Rough Sets and Fuzzy Sets Together. Intelligent Decision Support Kluwer Academic Publishers, Dordrecht. 203–232.

[6] Yeung D., Chen D., Tsang E.C., Lee J.W.T, Xizhao W. 2005. On the Generalization of Fuzzy Rough Sets. IEEE Transactıons on Fuzzy Systems. 13:343-361.

[7] Jensen, R. and Shen, Q. 2002. Fuzzy-rough sets for descriptive dimensionality reduction. In: Proceedings of IEEE International Conference on Fuzzy Systems, FUZZ-IEEE_02, May 12–17. 29–34.

[8] Jensen, R. and Shen, Q. 2004. Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough Based Approaches. IEEE Transaction Knowledge Data Engineering. 16:1457–1471.

[9] Jensen, R. and Shen, Q. 2007. Fuzzy-Rough Sets Assisted Attribute Selection. IEEE Transactions on Fuzzy Systems.15:73–89.

[10] Salido, F., Murakami, S. 2003. Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations. Fuzzy Sets and Systems. 139:635–660.

[11] Tsang, G.C.Y., Chen, D., Tsang, E.C.C., Lee, J.W.T. and Yeung, D.S. 2005. On attributes reduction with fuzzy rough sets. Proc. 2005 IEEE International Conference on Systems, Man and Cybernetics. 3: 2775–2780.

[12] Aydogan E.K., Gencer C. and AĞIRGÜN B. 2010. An Attribute Reduction Algorithm Based on Fuzzy Rough Set. 24th mini EURO Conference MEC EurOPT 2010.

[13] Tsang E.C.C., Chen D.G., Yeung D.S., Wang X.Z., Lee J.W.T. 2008. Attribute reduction using fuzzy rough sets. IEEE Trans. on Fuzzy. System.16:1130-1142.

[14] Wang R., Miao D., Hu G. 2006. Discernibility Matrix Based Algorithm for Reduction of Attributes. IEEE Computer Society, Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, (WI-IAT 2006 Workshops) (WI-IATW'06).477-480.

[15] Langley, P., Iba, W., Thomas K., 1994. An analysis of Bayesian Classifiers. In Proceedings of the tenth National conference of Artificial Intelligence, AAAI Press. 223-228.

[16] UCI (University of California at Irvine) Machine Learning Repository http://www.ics.uci.edu/~mlearn/MLRepository.html

[17] Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I.H. 2009. The WEKA Data Mining Software: An Update; SIGKDD Explorations.11:10-18.

[18] ROSETTA, a rough set toolkit for analyzing data, http://www.idi.ntnu.no/˜aleks/rosetta.