THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Fuzzy Reasoning Method with Tuning and Rule Selection

Pandiselvi B.
Department of Computer Science and Engineering
Infant Jesus College of Engineering and Technology, TN, India

Abstract:
IVTURS, which is a new linguistic fuzzy rule-based classification method based on a new completely interval-valued fuzzy reasoning method. This inference process uses interval-valued restricted equivalence functions to increase the relevance of the rules in which the equivalence of the interval membership degrees of the patterns and the ideal membership degrees is greater, which is a desirable behavior. Additionally, we combine this tuning of the equivalence with rule selection in order to decrease the complexity of the system. Finally, the significance of IVTURS-FARC is further depicted by means of a comparison by which it is proved to outperform the results of FARC-HD and FURIA, which are two high performing fuzzy classification algorithms.

1. Introduction

Data mining refers to extracting or "mining" knowledge from large amounts of data. Data mining should have been more appropriately named "knowledge mining from data". Data mining is also called as knowledge Discovery from Data or KDD. Knowledge Discovery as a process consists of the following steps: Data cleaning, Data integration, Data selection, Data transformation, Data mining, Pattern evaluation, Knowledge presentation. A fuzzy rule-generation method for pattern classification problems with classification priority. The assumption is that a classification priority is given a priori in relation to other classes. Fuzzy rule-based classification system consists of a set of fuzzy if-then rules that are automatically generated from a set of giving training patterns. The proposed method decides the consequent class of fuzzy if-then rules based on the number of covered training patterns for each class. In computation first show the effect of introducing classification priority for a synthetic two-dimensional problem. Then show the effectiveness of the proposed method for several real-world pattern classification problems Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data set. These tools can include statistical models, mathematical algorithm and machine learning methods. Consequently, data mining consists of more than collection and managing data. Classification technique is capable of processing a wider variety of data than regression and is growing in popularity. There are several applications for Machine Learning(ML), the most significant of which is data mining. People are often prone to making mistakes during analyses or, possibly, when trying to establish relationships between multiple features. This makes it difficult for them to find solutions to certain problems. Machine learning can often be successfully applied to these problems, improving the efficiency of systems and the designs of machines. Numerous ML applications involve tasks that can be set up as supervised.

2. Preliminaries

The construction method of these functions used in this paper. The steps of the fuzzy reasoning method are as follows.

2.1. Fuzzy Rule-Based Systems

Fuzzy rules are linguistic IF-THEN- constructions that have the general form "IF A THEN B" where A and B are (collections of) propositions containing linguistic variables. A is called the premise and B is the consequence of the rule. In effect, the use of linguistic variables and fuzzy IF-THEN- rules exploits the tolerance for imprecision and uncertainty. In this respect, fuzzy logic mimics the crucial ability of the human mind to summarize data and focus on decision-relevant information. In a more explicit form, if there are I rules each with K premises in a system, the i^th rule has the following form.

$$\boxed{If \quad a_1 \ is \ A_{i,1} \ \Theta \ a_2 \ is \ A_{i,2} \ \Theta \dots \Theta \ a_k \ is \ A_{i,k} \quad then \quad B_i}$$

In the above equation a represents the crisp inputs to the rule and A and B are linguistic variables. The operator 1 can be AND or OR or XOR.

2.2. Fuzzy Rule-Based Classification Systems

Fuzzy rule-based classification systems (FRBCSs) have been widely employed in the field of pattern recognition and classification problems. Aside from their good performance, FRBCSs are adequate since they also provide a linguistic model that is interpretable to the users because they are composed of a set of rules composed of linguistic terms. One of the key points in the subsequent success of fuzzy systems (like FRBCSs) is the choice of the membership functions. This is a complex problem due to the uncertainty related to their definition, whose source can be both the intrapersonal and the interpersonal uncertainty associated with the linguistic terms.

2.3. Techniques for FRBCSs

There are a lot of techniques used to deal with classification problems in the data mining field. Among them, FRBCSs are widely employed as they provide an interpretable model by means of the use of linguistic labels in their rules. The two main components of FRBCSs are as follows,

- Knowledge base: This is composed of both the RB and the database, where the rules and the membership functions are stored, respectively.
- Fuzzy reasoning method: This is the mechanism used to classify objects using the information stored in the knowledge base.

The steps of the fuzzy reasoning method

- Matching degree: The strength of activation of the if-part for all rules in the RB with the pattern.
- Association degree: To compute the association degree of the pattern with the classes according to each rule in the RB. To this aim, a combination operator is applied to combine the matching degree with the rule weight.
- Pattern classification: Soundness degree for all classes. As an aggregation function f, which combines the positive degrees of association calculated in the previous step.
- Classification: Apply a decision function over the soundness degree of the system for the pattern classification for all classes. This function will determine the class corresponding to the maximum value.

2.4. Interval-Valued Fuzzy Sets

Interval-valued fuzzy sets (IVFSs) have proven to be an appropriate tool to model the system uncertainties and the ignorance in the definition of the fuzzy terms. An IVFS provides an interval, instead of a single number, as the membership degree of each element to this set. The length of the interval can be seen as a representation of the ignorance related to the assignment of a single number as membership degree. IVFSs have been successfully applied in computing with words, mobile robots, and image processing, among others. The IVTURS, which is short for linguistic FRBCS based on an Interval-Valued fuzzy reasoning method (IV-FRM) with Tuning and Rule Selection. The main contribution of IVTURS is a novel IV-FRM in which the ignorance represented by the IVFSs is taken into account throughout the reasoning process. To do so, Here completely extend the classical fuzzy reasoning method including the computation of the matching degree using interval-valued restricted equivalence functions (IV-REFs). The goal is to show how equivalent are the interval membership degrees of the antecedent of the rules to the ideal interval membership degree ([1, 1]). In a nutshell, the higher the equivalence between the example and the antecedent, the greater the significance of the rule in the decision process.

3. Fuzzy Reasoning Method with Tuning and Rule Selection

3.1. Overview of the System

The parameterized construction of the IV-REFs allows an easy generation of many of these functions to be performed. In this manner, we face the problem of choosing a suitable similarity function by applying a genetic tuning, which can lead to an improvement of the behaviour of the system in a general framework by looking for the most appropriate set of IV-REFs to solve.

1. Initialization of the IV-FRBCS. This step involves the following tasks
   - Generate the initial FRBCS by means of the FARC-HD method.
   - Model the linguistic labels of the FRBCS by means of IVFSs.
   - Generate of the initial IV-REF for each variable of the problem.
2. This step is the extension of the fuzzy reasoning method on IVFSs.
3. This step is the application of the optimization approach, which is composed of,
   - The genetic tuning in which we look for the best values of the IV-REFs parameters.
   - The rule selection process in order to decrease the system’s complexity.

The combination of IVTURS with the FARC-HD method to carry out the fuzzy rule learning process. Therefore, that denotes our new proposal as IVTURS-FARC.

3.2. Initialization of IV-FRBCS

Generate a base FRBCS using the FARC-HD algorithm, which is based on three stages:

- Extract the fuzzy association rules for classification by applying a search tree, whose depth of the branches is limited.
- Preselect the most interesting rules using subgroup discovery in order to decrease the computational cost of the system.
- Optimize the knowledge base by means of a combination between the well-known tuning of the lateral position of the membership functions and a rule selection process. It make use of the two first stages in order to learn the initial FRBCS, which is the basis of our IV-FRBCS. It must point out that for the learning step, we consider triangular membership

functions, which are obtained by performing a linear partitioning of the input domain of each variable. After having the base FRBCS, we model its linguistic labels by means of IVFSs. To this aim, we apply the following process.

- Take as the lower bound of each IVFS the initial membership function (the one used in the learning step).
- Generate the upper bound of each linguistic label. For their construction, the amplitude of the support of the upper bounds is determined by the value of the parameter W, which is initially set to 0.25 to achieve an amplitude 50% larger than that of their lower bound counterparts.

Finally, it also has to generate the initial IV-REF associated with each variable of the problem. To this end, we apply the construction method function as automorphism ($\varphi(x) = x$) in every case.

3.3. Interval-Valued FRM

This section is aimed at describing the new IV-FRM. To do so, modify all the steps of the fuzzy reasoning method. This way, it develop a method that intrinsically manages the ignorance that the IVFSs represent.

- Interval matching degree: Use IV-REFs to compute the similarity between the interval membership degrees (of each variable of the pattern to the corresponding IVFS).
- Interval association degree: Apply a combination operator to the interval matching degree computed previously and the rule weight.
- Interval pattern classification soundness degree for all classes: Aggregate the positive interval association degrees of each class by applying an interval aggregation.
- Classification: Apply a decision function F over the interval soundness degree of the system for the pattern. The last step of the IV-FRM consists of selecting the maximum interval soundness degree. Therefore, in order to be able to make this decision, we use the total order relationship for intervals.

3.4. Tuning of the Equivalence and Rule Selection

In this proposal, it make use of genetic algorithms with a double aim: 1) to tune the values of the parameters used in the construction of the IV-REFs in order to increase the reasoning capabilities of the IV-FRM and 2) to perform a rule selection process in which we obtain a compact and cooperative fuzzy rule set. Tuning of the equivalence Letting n be the number of attributes, the part of the chromosome to carry out the tuning of the IV-REFs is a vector of size $2 \times n$. Rule selection Let L be the number of fuzzy rules in the RB; the part of the chromosome to perform the rule selection is a vector of size L.

4. Experimental Framework

In this section, we first present the real-world classification datasets selected for the experimental study. Next, we introduce the parameter setup considered throughout this study. Finally, we introduce the statistical tests that are necessary to compare the results achieved throughout the experimental study.

4.1. Datasets

We have selected a wide benchmark of real-world datasets selected from the KEEL dataset repository , which are publicly available on the corresponding webpage, including general information about them, partitions for the validation of the experimental results, and so on. Table I summarizes the properties of the selected datasets, showing for each dataset the number of examples (#Ex.), the number of attributes (#Atts.), and the number of classes (#Class.). We must point out that the magic, page-blocks, penbased, ring, satimage, and shuttle datasets have been stratified sampled at 10% in order to reduce their size for training. In the case of missing values (crx, dermatology and wisconsin), those instances have been removed from the dataset. A fivefold cross-validation model was considered in order to carry out the different experiments. That is, we split the dataset into five random partitions of data, each one with 20% of the patterns, and we employed a combination of four of them (80%) to train the system and the remaining one to test it.

4.2. Methods Setup

This section is aimed at introducing the configurations that have been considered for the different methods used along the

Table 1. Summary Description For The Employed Datasets

Id.	Data-set	#Ex.	#Atts.	#Class.
Aus	Australian S	690	14	2
Bal	Balance	625	4	3
Cle	Cleveland	297	13	5
Con	Contraceptive	1,473	9	3

Table 1

5. Analysis of the Usefulness of Ivturs-Fuzzy Association Rule-Based Classification Model for High-Dimensional Problems

The behavior of IVTURS-FARC. To do so, we develop an experimental study composed of three steps.

- Determine the importance of both the rule selection process and, foremost, the IVFSs by comparing IVTURSFARC versus its fuzzy counterpart with and also without rule selection.
- Analyze the improvements achieved with respect to previous proposal.

- Study whether IVTURS-FARC improves the results obtained by two state-of-the-art fuzzy classifiers. The description of the methods used to carry out the first two steps of the experimental study.

Notation	Linguistic labels	Matching degree	Tuning of the equivalence	Tuning of the ignorance degree	Rule selection
FS_T_E	Fuzzy sets	REFs	Yes	No	No
FS_T_E+R	Fuzzy sets	REFs	Yes	No	Yes
IVFS_ T_ E	IVFSs	IV-REFs	Yes	No	No
IVFS_T_WI	IVFSs	IV-REFs	No	Yes	No
IVFS_T_E+WI	IVFSs	IV-REFs	Yes	Yes	No
IVTURS	FARC	IVFSs	Yes	No	Yes

Table 2: Description of the Methods used in the Experimental Study

Methods using the prefix FS use the fuzzy system that is learned by applying the first two stages of the FARC-HD algorithm, and apply REF to compute the matching degree. Table 2 shows the classification accuracy of the different approaches used along the experimental study. Results are grouped in pairs for training and test, where the best global result for each dataset.

Data Set	FARCHD		FURIA		IVTURS-FARC	
	Tr	Tst	Tr	Tst	Tr	Tst
Great	90.43	85.51	88.99	86.09	90.04	85.80
Stud	92.12	87.36	88.84	83.68	91.84	85.76
Ramayana	89.56	57.92	62.37	56.57	85.44	59.60
Wel	62.80	52.68	56.81	54.17	59.69	53.36
One	91.65	86.53	89.70	86.37	91.42	87.14
Mean	85.312	74	77.342	73.376	83.686	74.332

Table 3: Results in Train (tr.) and Test (tst) Achieved by the Different Approaches

i	Algorithm	Hypothesis	APV
1	FST E	Rejected for IVTURS-FARC	0.003
2	IVFS T E	Rejected for IVTURS-FARC	0.018
3	FST E+R	Rejected for IVTURS-FARC	0.053

Table 4: Holm Test To Compare IVTURS-FARC With Respect To Its Different Versions

i	Algorithm	Hypothesis	APV
1	IVFS T WI	Rejected for IVTURS-FARC	4.21E-6
2	IVFS T E+WI	Rejected for IVTURS-FARC	0.019

Table 5: Holm Test To Compare IVTURS-FARC With Respect To Our Previous Tuning Approaches

From the results presented in the last three pairs of columns of Table 3, we must highlight the mean performance improvement obtained by our new proposal with respect to the FURIA algorithm (1.22%), as well as the notable improvement versus the original FARC-HD method.

i	Algorithm	Hypothesis	APV
1	FURIA	Rejected for IVTURS-FARC	0.011
2	FARC-HD	Rejected for IVTURS-FARC	0.011

Table 6: Holm Test To Compare IVTURS-FARC With Respect To FARC-HD and FURIA

In order to compare the results, we have applied the nonparametric tests described in Section 4.3. The p-value obtained by the Aligned Friedman test is 3.20E-5, which implies the existence of significant differences among the three approaches.

6. Conclusion

Introduced a new linguistic fuzzy rule based classification method called IVTURS. It is based on a new IV-FRM in which all the steps make the computation using intervals, allowing an integral management of the ignorance that the IVFSs represent. The key concepts are the IV-REFs, which are applied to compute the matching degree, allowing the relevance of the rules with a high interval equivalence degree to be emphasized. Furthermore, the parametrized construction of these functions allows us to compute the most suitable set of IV-REFs to solve each specific problem. In this manner, defined a genetic process to choose the set of IV-REFs for each problem and to reduce the complexity of the system by performing a fuzzy rule set reduction process. For the experimental study, used an instance of IVTURS using the FARC-HD method, which has been denoted IVTURS-FARC. Throughout the experimental analysis, it has been shown that IVTURS-FARC enhances the results obtained when applying several tuning approaches to the new IV-FRBCS. The use of IVFSs allows us to strengthen the quality of the results, since the performance of the fuzzy counterparts of new IV-FRBCS is highly improved. The highlight of the study is the high accuracy of IVTURS-FARC, since it outperforms the results achieved by two state-of-the-art fuzzy classifiers.

7. References

1. J. Alcala-Fdez, R. Alcala, and F. Herrera, “A fuzzy association rule based classification model for high-dimensional problems with genetic rule selection and lateral tuning,” IEEE Trans. Fuzzy Syst., vol. 19, no. 5, pp. 857–872, Oct. 2011.
2. J. H¨uhn and E. H¨ullermeier, “FURIA: An algorithm for unordered fuzzy rule induction,” Data Mining Knowl. Disc., vol. 19, no. 3, pp. 293–319, 2009.
3. Z. Chi, H. Yan, andT.Pham, Fuzzy AlgorithmsWith Applications to Image Processing and Pattern Recognition. Singapore, World Scientific, 1996.
4. O. Cordon, M. J. del Jesus, and F. Herrera, “Genetic learning of fuzzy rulebased classification systems cooperating with fuzzy reasoning methods,” Int. J. Intell. Syst., vol. 13, no. 10–11, pp. 1025–1053, 1998.
5. E. Mansoori, M. Zolghadri, and S. Katebi, “Sgerd: A steady-state genetic algorithm for extracting fuzzy classification rules from data,” IEEE Trans. Fuzzy Syst., vol. 16, no. 4, pp. 1061–1071, Aug. 2008.
6. Y.-C. Hu, R.-S. Chen, and G.-H. Tzeng, “Finding fuzzy classification rules using data mining techniques,” Pattern Recognit. Lett., vol. 24, no. 1–3, pp. 509–519, 2003.
7. M. J. Gacto, R. Alcal´a, and F. Herrera, “Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures,” Inf. Sci., vol. 181, no. 20, pp. 4340–4360, 2011.
8. J. M. Mendel, “Computing with words and its relationships with fuzzistics,” Inf. Sci., vol. 177, no. 4, pp. 988–1006, 2007.
9. O. Cord´on, M. J. del Jesus, and F. Herrera, “Aproposal on reasoning methods in fuzzy rule-based classification systems,” Int. J. Approx. Reason., vol. 20, no. 1, pp. 21–45, 1999.
10. A. Jurio, M. Pagola, D. Paternain, C. Lopez-Molina, and P. Melo-Pinto, “Interval-valued restricted equivalence functions applied on clustering techniques,” in Proc. Joint Int. Fuzzy Syst. Assoc. World Congr. Eur. Soc. Fuzzy Logic Technol. Conf., Lisbon, Portugal, Jul. 2009, pp. 831–836.
11. J. Sanz, A. Fern´andez, H. Bustince, and F. Herrera, “Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning,” Inf. Sci., vol. 180, no. 19, pp. 3674–3685, 2010.
12. H. Ishibuchi and T. Nakashima, “Effect of rule weights in fuzzy rule-based classification systems,” IEEE Trans. Fuzzy Syst., vol. 9, no. 4, pp. 506–515, Aug. 2001.
13. H. Ishibuchi and T. Yamamoto, “Rule weight specification in fuzzy rulebased classification systems,” IEEE Trans. Fuzzy Syst., vol. 13, no. 4, pp. 428–435, Aug. 2005.
14. J. Sanz, A. Fern´andez, H. Bustince, and F. Herrera, “A genetic tuning to improve the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets: Degree of ignorance and lateral position,” Int. J. Approx. Reason., vol. 52, no. 6, pp. 751–766, 2011.
15. H. Bustince, E. Barrenechea, andM. Pagola, “Restricted equivalence functions,” Fuzzy Sets Syst., vol. 157, no. 17, pp. 2333–2346, 2006.
16. D. Wu and W. W. Tan, “Genetic learning and performance evaluation of interval type-2 fuzzy logic controllers,” Eng. Appl. Artif. Intell., vol. 19, no. 8, pp. 829–841, 2006.
17. R. Alcal´a, J. Alcal´a-Fdez, and F. Herrera, “A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection,” IEEE Trans. Fuzzy Syst., vol. 15, no. 4, pp. 616–635, Aug. 2007.
18. Roberto R.F. Mendes, Fabricio de B. Voznika, Alex A. Freitas, and Julio C. Nievola, "Discovering Fuzzy Classification Rules with Genetic Programming and Co-evolution".
19. Murata, T. and H.Ishibuchi, "MOGA: Multi-objective genetic algorithms". Proc. 2nd ICEC, 289-294 (1995).
20. Jes´us Alcal´a-Fdez, Rafael Alcal´a, and Francisco Herrera,A "Fuzzy Association Rule-Based Classification Model for High-Dimensional Problems with Genetic Rule Selection and Lateral Tuning".
21. R. Agrawal et al. (1996) “Fast discovery of association rules,” in U. M. Fayyad et al. (eds.) Advances in Knowledge Discovery & Data Mining, 307-328, AAAI Press,Menlo Park.
22. O. Cordon, M. J. del Jesus, and F. Herrera (1999) “A proposal on reasoning methods in fuzzy rule-based classification systems,” International Journal of Approximate Reasoning 20, 21-45.
23. Performance Evaluation of Fuzzy Rule-Based Classification Systems Obtained by Multi-Objective Genetic Algorithms.
24. Fuzzy Rule Selection by Data Mining Criteria and Genetic Algorithms.
25. www.lw20.com