A Fuzzy Rule Base Minimization Perspective in XAI

F. Camastra[1], A. Ciaramella[1], S. Sposato[1] and A. Staiano[1]

[1]*Department of Science and Technology, University of Naples Parthenope, Centro Direzionale Isola C4, I-80143, Napoli, Italy*

Abstract

Fuzzy rule-based systems are raising great interest in the last years in eXpalianable Artificial Intelligence. These systems represents knowledge easily understood by humans but they are not interpretable *per se*. They, in fact, must remain simple and understandable, and the rule base must be compactness. In this work a fuzzy rule base minimization approach based on rough sets theory and a greedy algorithm is proposed. The reduction of the fuzzy rules makes the rule base simpler, and thus easier to produce explainable inference systems (e.g., decision support systems and recommenders). Encouraging results are obtained validating and comparing the methodology on data of UCI benchmark.

Keywords

Fuzzy rule base, Explainable Artificial Intelligence, Rough sets, Greedy algorithms.

1. Introduction

In recent years, Computational Intelligence methodologies are experiencing considerable interest in eXplainable Artificial Iintelligent (XAI) [1, 2]. Techniques for XAI can be model agnostic (i.e. they can be applied to any AI algorithm), or model specific (i.e. they can be only applied to a specific AI algorithm). Moreover, they can be ante-hoc (transparent or "white box/glass box" approaches explainable by design or inherently explainable) or post-hoc (divided into global explanations or local explanations) explainability methods [3]. Ante-hoc methods are explainable by design or inherently explainable methods and are also referred to as transparent approaches, which are model specific, include linear and logistic regression, decision trees, k-nearest neighbors, fuzzy inference systems, rule-based learners, general additive models, and Bayesian models. Fuzzy Rule-based Systems (FRSs), are raising great interest in XAI in the last years as ante-hoc methodologies [3]. The main components of any FRS are the knowledge base (KB) and the inference engine module. The KB comprises all the fuzzy rules within a rule base (RB), and the definition of the fuzzy sets in the data base. The inference engine includes a fuzzification interface, an inference system, and a defuzzification interface [4, 5]. However, it must be emphasized that FRSs must remain simple and understandable, since they are not interpretable *per se* [6, 2]. It is important to take account of different issues in order to

obtain FRBSs that represent knowledge easily understood by humans. Among others, the rule base compactness or the semantic comprehensibility of the fuzzy partitions must be stressed [7, 6]. Moreover, the EFSs must be properly designed to obtain the desired trade-off between accuracy and explainability for the problem at hand. In this work we introduce a fuzzy rule base minimization approach based on rough sets theory and a greedy based approach.

The paper is organized as follows. In Section 2 we present the proposed methodology and the used methods. Furthermore, in Section 3 the results of experiments on benchmark data are presented. Finally, the authors draw conclusions in Section 4.

2. Reducing Rules Approach

The problem of finding the useless attributes for the correct classification is intractable when the number of attributes is large and algorithms that provide suboptimal approximated solutions must be explored. The *Reducing Rules and Conditions* (*RRC*) algorithm provides a greedy approximated solution to the problem of identifying and deleting in a set of fuzzy rules the attributes that are irrelevant for the correct classification. RRC approach is composed by five steps as described in Figure 1. In order to evaluate a rule pattern in the algorithm search, a property, i.e., *efficiency*, is associated to each rule pattern (see [8] for details). In particular, the efficiency definition reflects the main goal of RRC algorithm, i.e., searching for each consequent the rule patterns with the smallest antecedent length that cover the largest number of rules, having, at the same time, the minimal overlap with the other consequents.

2.1. Building decision tables from fuzzy rules

In this stage each rule of the a fuzzy knowledge base, $\mathcal{F}$ is represented into a decision system model, labeling antecedents and consequents of the fuzzy rule as the condition and decision attributes, respectively. Firstly, a label is associated to each pair (Linguistic Variable, Value) and then, each pair in fuzzy rules is replaced with the respective labels (e.g., High $\rightarrow$ H).

2.2. Sorting attributes by their significance

In this phase, the relevance for each attribute (i.e., for each fuzzy relation) is computed. To this purpose the attributes relevance is measured by its *significance* by using rough sets theory [9]. Given two sets of attributes P and R, the significance of an attribute x [10], denoted by σ_x^R, is defined as follows:

$$s_x^R(P) = \gamma_R(P) - \gamma_{R-\{x\}}(P) \tag{1}$$

where the parameter $\gamma_R(P)$, that always takes values between 0 and 1, represents the fraction of the objects that can be classified correctly [9]. It is worthwhile to remark that the so-defined significance is relative since it depends both on P and R sets.

2.3. Building prefix tree

The construction of the prefix tree takes processing one rule at a time. In particular, for each attribute, a node is generated, which becomes the child of the node of the previous attribute
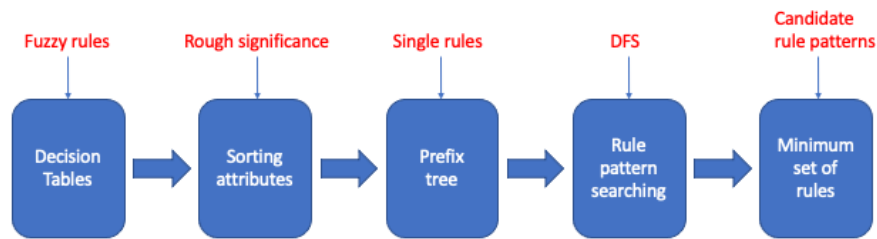
Figure 1: Main steps of the fuzzy rule minimization approach.

and the parent of the node of the next attribute. The first attribute, becomes the child node of the root tree, which, not having any attribute, is an empty node. The tree has the property that all the descendants of a node shares the prefix associated with the node. This implies that two rules with the same initial condition share the same node associated with that condition.

2.4. Rule pattern searching

The search algorithm uses the prefix tree, constructed in the previous stage, as a decision tree and carries out a left-to-right depth-first search (DFS) visit of the tree for each decision [11]. For each iteration, a reduction of the complete tree is computed and then, the search algorithm is performed. The number of iterations is fixed by a parameter. The reduction of the complete tree is obtained simply by randomly discarding some attributes from decision table and constructing, consequently pruned prefix tree.

2.5. Sorting rule patterns and building the minimum set of rules

The final result of rule pattern searching is a list of candidate rule patterns which is usually oversized compared with the initial knowledge base. Therefore, in these stage it is extracted only a subset of candidate rule patterns, that can cover all the decisions produced by the initial knowledge base and, at the same time, minimize the number of rules and conditions. The set of first rule patterns that cover all the decisions produced by the initial knowledge base.

3. Experimental Results

The proposed algorithm has been validated on two UCI benchmarks, i.e., *mushroom* [12], *breast-cancer* [13] [1], and was compared with Ripper [14], Part [15], Lem2 [16] algorithms. Benchmark characteristics are summarized in Table 1. The performance of the algorithms have been measured in terms of number of rules extracted, *Coverage*, *Accuracy*[2]. In Tables 2 and 3 we describe the results obtained on mushroom and breast cancer benchmarks, respectively. We observe that the proposed methodology permits to obtain a low number of rules and attributes maintaining high accuracy and coverage.

benchmark	number of rules	number of attributes
Mushroom	8214	178728
Breast-Cancer	286	2574

Table 1
Benchmark characteristics.

benchmark	number of rules	number of attributes	Accuracy	Coverage
Ripper	8	12	100%	100%
Part	13	20	100%	100%
Lem2	8	78	100%	100%
Genetic Algorithm	2769	4325	100%	98.7%
RRC	10	29	100%	100%

Table 2
Results on mushroom benchmark.

benchmark	number of rules	number of attributes	Accuracy	Coverage
Ripper	17	56	90.2%	100%
Part	54	149	91.9%	100%
Lem2	124	574	98.5%	100%
Genetic Algorithm	1042	3325	98.6%	98.7%
RRC	53	238	100%	100%

Table 3
Results on breast cancer benchmark.

---

[1]informations about the other benchmarks can be found on `http:\\ archive.ics.uci.edu/ml/datasets.php`

[2]Coverage: percentage of the rules of initial knowledge base that are covered by the set of rules generated by compressing algorithm; Accuracy: percentage of the correct classification of the set of the rules generated by the compressing algorithm.

4. Conclusion

In this work a fuzzy rule base minimization approach based on rough sets theory and a greedy algorithm has been introduced. Rough sets theory has been used for ordering significance of the attributes and greedy approach for building a prefix tree used as a decision tree carrying out left-to-right depth-first search visit. The proposed approach is validated and compared on UCI benchmarks resulting encouraging results. In the next future the authors will focus on the theoretic background of the methodology (e.g., efficiency measures and optimal path demonstration). Further experiments and comparisons will be conducted on different data and for addressing the explainability of the obtained fuzzy rules.

Acknowledgments

References

[1] C. Mencar, J. Alonso, Paving the way to explainable artificial intelligence with fuzzy modeling, volume 24, 2018, pp. 215–227.

[2] A. Fernandez, F. Herrera, O. Cordon, M. del Jesus, F. Marcelloni, Evolutionary fuzzy systems for explainable artificial intelligence: Why, when, what for, and where to?, IEEE Computational intelligence magazine 14 (2019) 69–81.

[3] A. Knapič, A. Malhi, R. Saluja, K. Främling, xplainable artificial intelligence for human decision-support system in medical domain, arXiv (2021).

[4] A. Ciaramella, R. Tagliaferri, W. Pedrycz, A. Di Nola, Fuzzy relational neural network, International Journal of Approximate Reasoning 41 (2006) 146–163.

[5] F. Camastra, A. Ciaramella, V. Giovannelli, M. Lener, V. Rastelli, A. Staiano, G. Staiano, A. Starace, A fuzzy decision system for genetically modified plant environmental risk assessment using mamdani inference, Expert Systems with Applications 42 (2015) 1710–1716.

[6] J. M. Mendel, P. P. Bonissone, Critical thinking about explainable ai (xai) for rule-based fuzzy systems, IEEE Transactions on Fuzzy Systems 14 (2019) 69–81.

[7] L. Jara, A. González, R. Pérez, A preliminary study to apply the quine mccluskey algorithm for fuzzy rule base minimization, 2020, pp. 1–6.

[8] T. Hastie, R. Tibshirani, R. Friedman, The Elements of Statistical Learning, Springer, New York, 2009.

[9] Z. Pawlak, Rough sets, Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, 1991.

[10] M. Modrzejewski, Feature selection using rough sets theory, in: Machine Learning: ECML-93, Springer, 1993, pp. 213–226.

[11] T. Cormen, C. Leiserson, R. Rivest, Introduction to Algorithms, MIT Press, 2009.

[12] J. Schlimmer, Concept Acquisition through representation adjustment, Technical Report, University of California, Irvine, PhD Thesis, 1987.

[13] R. Michalski, I. Mozetic, J. Hong, J. N. Lavrac, The multi-purpose incremental learning system aq15 and its testing application to three medical domains, in: Proceedings of the Fifth National Conference on Artificial Intelligence, Morgan Kaufman, 1986, pp. 1041–1045.

[14] W. Cohen, Fast effective rule induction, in: Machine Learning Proceedings 1995, Morgan Kauffman, 1995, pp. 115–123.

[15] E. Frank, I. Witten, Generating accurate rulesets without global optimization, in: 1998 Working Papers, University of Waikato, Department of Computer Science, 1998, pp. 1–15.

[16] J. Grzymala-Busse, Lers-a system for learning from examples based on rough sets, in: Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory, Dordrecht: Springer Netherlands, 1992, pp. 3–18.