Information Sciences 349–350 (2016) 199–215

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Genetic generation of fuzzy systems with rule extraction using formal concept analysis

M.E. Cintra [a,*], H.A. Camargo [b], M.C. Monard [c]

[a] Federal Territorial University of the Semi-Arid, Mossoro 59621400, RN Brazil
[b] Federal University of São Carlos (UFSCar), São Carlos, Brazil
[c] University of São Paulo (USP), São Paulo, Brazil

ARTICLE INFO

Article history:
Received 19 January 2015
Revised 13 January 2016
Accepted 12 February 2016
Available online 18 February 2016

Keywords:
Genetic fuzzy systems
Formal concept analysis
Classification

ABSTRACT

Fuzzy classification systems have been widely researched with many approaches proposed in the literature. Several methods are available for the automatic definition of fuzzy classification systems, which basically comprehend two tasks: i) the definition of the attributes in terms of fuzzy sets, and ii) the generation of a rule set containing the domain knowledge, named fuzzy rule base. Genetic Fuzzy Systems are used to learn or tune in fuzzy classification systems. Some genetic approaches for learning the fuzzy rule base require the previous extraction of a set of rules to be used as the genetic search space. In this paper, we present the FCA-Based method, a proposal for the automatic generation of fuzzy rule bases, which extracts a set of rules using the formal concept analysis theory directly from data. After extracting the rules forming the genetic search space, FCA-Based uses a genetic algorithm to select the final rule base. The last step of the FCA-Based method is a rule pruning step in order to improve the interpretability of the fuzzy rule bases. The extraction of rules proposed for the FCA-Based algorithm presents polynomial complexity and does not require the predefinition of the number of rules to be extracted. As it extracts rules directly from data, the proposed method avoids the random extraction of rules. It also presents the advantage of automatically extracting rules with variable number of conditions in their antecedents. A feature subset selection method, specifically designed for fuzzy classification systems, is integrated into the FCA-Based method in order to reduce the search space of solutions. The FCA-Based method is detailed and compared to eight different rule-based fuzzy systems. Experimental results using 27 benchmark datasets and a 10-fold cross-validation strategy show that FCA-Based presents higher accuracy and statistically significant difference with seven of the eight compared methods.

1. Introduction

The task of classification has been widely researched by both, the machine learning and soft computing communities. The methods proposed in the literature for the task of classification include different approaches, such as artificial neural networks [41], support vector machines [51], as well as rule-based approaches, including decision trees [17,21,48] and Genetic Fuzzy Systems (GFS) [6,7,10,20,23,28,45], among others. Special attention has been given to the GFSs with a large

* Corresponding author. Tel.: +55 8430616039.
E-mail addresses: mecintra@gmail.com (M.E. Cintra), heloisa@dc.ufscar.br (H.A. Camargo), mcmonard@icmc.usp.br (M.C. Monard).

http://dx.doi.org/10.1016/j.ins.2016.02.026
0020-0255/

number of proposals found in the literature [19]. GFSs combine the power of the global search of Genetic Algorithms (GA) with the fuzzy theory to provide accurate and interpretable rule-based systems. In [32], Prof. Herrera presents a classification of the various genetic fuzzy methods found in the literature according to a comprehensive taxonomy, while in [52], Shukla and Tripathi present a survey on genetic fuzzy systems, focusing on the interpretability $\times$ accuracy trade-off. In [27], the authors present a taxonomy of evolutionary fuzzy systems, as well as applications, new trends and challenges. A special type of Genetic Fuzzy Systems, which has been widely researched recently, is the Multi Objective Evolutionary Fuzzy Systems (MOEFS). In this type of systems, a set of solutions with different degrees of accuracy and interpretability is generated, by means of Multi Objective Genetic Algorithms (MOGA)) [8,13,25,26,36,49].

In this work, we focus on the genetic learning of Fuzzy Rule Bases (FRBs) using a predefined Fuzzy Data Base (FDB) and previous extraction of candidate rules. This type of approach is classified as *genetic rule selection with a priori rule extraction* [32]. The adoption of this approach is based on its advantages for our proposal: by using a predefined FDB, the GA focus its search exclusively on the generation of suitable FRBs. Also, by using a predefined FDB, it is possible to automatically build and extract rules, *i.e.*, a set of candidate rules that define the genetic search space. Another advantage of this approach is the fact that, since the search space is defined before the GA search process, it is possible to use simple and efficient codifications for the chromosomes that improve the use of genetic operators and accelerate the whole genetic process.

Regarding the definition of the FDB, many approaches found in the literature are based on GAs [3,23,28,45,47], the concept of fuzzymetric arcs [39], clonal selection algorithms [1], artificial neural networks [9], fuzzy clustering algorithms [42], as well as the use of special indexes, such as fuzzy entropy [2], the 2-tuple linguistic membership representation [28,33,44], and methods based on interval-valued fuzzy reasoning methods [50], among others. However, the equalized universe method [14], which uses fixed numbers of fuzzy sets evenly distributed for all attributes, has been widely adopted due to its simplicity and the interpretability of the generated partitions. In order to provide a fast and robust approach to define FDBs, we proposed the FuzzyDBD method in [18], which combines an attribute fuzzification step with the use of an estimation function to estimate the number of fuzzy sets for each attribute independently.

Regarding the generation of FRBs, as previously stated, some genetic proposals require a previous extraction of a pool of rules that form a genetic search space, which is followed by a genetic rule selection step. The first approaches found in the literature for this task performs the pre-selection of rules according to certain measures, such as the confidence and support of the rules [37], or by limiting the total number of conditions of the extracted rules [7]. A similar approach, using fuzzy association rules is proposed in [8], which is based on two steps: i) a fuzzy rule extraction based on association rules, and ii) a rule prescreening to reduce the number of extracted rules. In [35], Ishibuchi & Murata propose the extraction of a random set of antecedent parts of rules, whose consequents, *i.e.*, the class, are determined by calculating the degree of confidence of all possible classes and assigning the class with highest degree of confidence for each antecedent, forming a rule. In [28], similarly to what is proposed in [35], a set of rules is generated taking into account all possible combinations of the values of input attributes, considering from two to seven linguistic values for each attribute. This set is used to create the initial population of a GA, which is complete with randomly generated rules.

Another related approach, named DoC-Based, performs the pre-selection of rules using their degree of coverage [15], but in an alternative fashion: for datasets defined by a reasonably small number of attributes, the algorithm generates all possible rules, while for datasets described by many attributes, the algorithm generates a large set of random rules, then it calculates the degree of confidence of the generated rules and uses this degree of confidence to pre-select a subset of these rules to form the genetic search space. These previous approaches depend on a large number of rules, generated using the fuzzy linguistic values defining the attributes. The process of generating and evaluating a large number of rules poses an exponential complexity problem. Thus, depending on the number of attributes and linguistic values defining the attributes, a pre-selection of the most relevant attributes is essential.

In order to avoid the complexity problem of generating a large set of candidate rules, we propose the use of Formal Concept Analysis (FCA) to extract classification rules and form the genetic search space. The proposed method for rule selection is the main contribution of this work. The FCA theory on data analysis identifies conceptual structures in data sets. Experimental results showed that our proposal is able to extract classification rules with the following desirable characteristics:

- The rules are extracted directly from the data, avoiding the extraction of a potentially large number of unnecessary/useless rules;
- Unlike the exhaustive generation of rules, which has exponential complexity, by using FCA the complexity of the process is typically polynomial [43];
- The extracted rules present variable number of conditions in the antecedent, improving the interpretability of the generated FRBs.

In this paper, we detail the proposed FCA-Based method and compare it to eight different rule-based fuzzy classification proposals. The FCA-Based method shows higher accuracy than the compared methods. Moreover, there is significant difference between the FCA-Based method and seven of the eight compared methods.

The remainder of this paper is organized as follows: Section 2 discusses Fuzzy Classification Systems. Section 3 presents the basic concepts of FCA and our rule extraction method using FCA. Section 4 presents the FCA-Based method. Section 5 presents the experiments and Section 6 the conclusions and future work.