Granul. Comput. (2017) 2:1–12
DOI 10.1007/s41066-016-0022-5

ORIGINAL PAPER

The development of granular rule-based systems: a study in structural model compression

Sharifah Sakinah Syed Ahmad[1] · **Witold Pedrycz**[2]

Received: 20 April 2016 / Accepted: 1 June 2016 / Published online: 23 June 2016

Abstract In this study, we develop a comprehensive design process of granular fuzzy rule-based systems. These constructs arise as a result of a structural compression of fuzzy rule-based systems in which a subset of originally existing rules is retained. Because of the reduced subset of the originally existing rules, the remaining rules are made more abstract (general) by expressing their conditions in the form of granular fuzzy sets (such as interval-valued fuzzy sets, rough fuzzy sets, probabilistic fuzzy sets, etc.), hence the name of granular fuzzy rule-based systems emerging during the compression of the rule bases. The design of these systems dwells upon an important mechanism of allocation of information granularity using which the granular fuzzy rules are formed. The underlying optimization consists of two phases: structural (being of combinatorial character in which a subset of rules is selected) and parametric (when the conditions of the selected rules are made granular through an optimal allocation of information granularity). We implement the cooperative particle swarm optimization to solve optimization problem. A number of experimental studies are reported; those include fuzzy rule-based systems.

Keywords Rule-based systems · Structural compression · Optimal allocation of information granularity · Particle swarm optimization · Granular fuzzy sets

✉ Sharifah Sakinah Syed Ahmad
sakinah@utem.edu.my

[1] Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, 76100 Melaka, Malaysia

[2] Department of Electrical and Computer Engineering, University of Alberta, Edmonton T6G 2G7, Canada

1 Introduction

There have been a large number of studies and applications on fuzzy rule-based systems. The rules are viewed as descriptors of individual, local pieces of knowledge, especially when forming a global mapping from the space of conditions to the space of conclusions. When dealing with a large number of rules, emerges an interesting and practically viable question about a reduction of the number of rules, so that a small subset of the most representative rules can be formed (Gacto et al. 2011; Antonelli et al. 2016; Zhou and Gan 2008; Baranyi and Yam 2000; Riid and Rüstern 2011; Cordon 2011; Villar et al. 2012; Juang and Chen 2013). The practical relevance stems from the two facts. First, the smaller number of rules enhances their readability meaning that the transparency of the reduced model becomes enhanced. Second, computing overhead is reduced. Starting from the set of rules "if x is A_j, then y is B_j" $j = 1, 2, \ldots, N$, the reduction of the model leads to the subset of rules "if x is A_i, then y is B_i" $i = 1, 2, \ldots, I$ where $I \ll N$. Surprisingly, the reduced rules do not reflect a fact they are the subset of the original far larger collection of rules. Intuitively, we might have anticipated that the reduced rule set reflects the reduction aspect by having a level of abstraction of the fuzzy sets standing in the condition parts of the rules being elevated. In other words, the reduced set of rules comprises the conditional statements of the form "if x is $\boldsymbol{G}(A_i)$, then y is B_i". The increased level of abstraction (generality) is realized by forming a granular augmentation of the original fuzzy set A_j by generalizing it to the granular fuzzy set $\boldsymbol{G}(A_j)$ viz. an interval fuzzy sets, fuzzy set of type-2, shadowed fuzzy sets, probabilistic (fuzzy) sets, and other generalizations. In a nutshell, the term granular fuzzy set stands for the generalization of the fuzzy set in which the original numeric value of membership, say $A_j(x)$ at point "x" is generalized to the granular value (interval, fuzzy set in [0, 1], probability density function, etc.). This granular nature of the proposed construct is directly associated with the reduced number of rules to compensate for the reduction of the rule base, the rules have to be made more abstract.

Assuming that the reduced set of rules has been formed via the collection of rules "if x is $\boldsymbol{G}(A_j)$, then y is B_j" has been decided upon, the fundamental question arises as to the formation of the granular fuzzy sets. The underlying design principle is that of an optimal allocation of information granularity. The values of the membership grades are non-numeric, for example, intervals or membership functions. Given a certain predetermined level of information granularity α, we allocate it among the elements of the original fuzzy set (by making it granular), so that a balance of information granularity is met and a certain optimization criterion is maximized. The optimization criterion used to guide the process of granularity allocation expresses an extent to which the results of inference process realized with the use of all the rules are "covered" by the results formed by the reduced rule-based system.

The development of the granular rule-based system comprises two important and intertwined phases, namely, a selection of a subset of the rules and a formation of the granular rule-based system. Given the combinatorial character of the first phase and a nonlinear nature of the overall process of granularity allocation, in the study, we consider a particle swarm optimization environment (PSO) as well as its generalized cooperative version.

The study is organized as follows. In Sect. 2, we discuss the underlying concept. In the sequel, we discuss the designing process of the granular rules. A suite of protocols of allocation of information granularity is presented. In Sect. 4, we describe the PSO environment using which the granular fuzzy rule-based system is constructed. In Sect. 4, experimental studies are given. Finally, conclusions and some prospects of further research are presented in Sect. 5.

Regarding the notation, capital letters (A, B, A_i, etc.) are used to denote fuzzy sets defined in the discrete universes of discourse. The notation $\boldsymbol{G}(A)$ is reserved to describe the granular fuzzy set. Furthermore, we assume that the fuzzy sets A_i standing in the conditions of the rules have infinite supports.

2 From fuzzy rule-based models to granular fuzzy rule-based models: the concept

The essence of fuzzy rule-based systems is inherently associated with the inference schemes of approximate reasoning

$$\begin{array}{l} x \text{ is } A \\ \text{if } x \text{ is } A_{\mathrm{i}} \text{ then } y \text{ is } B_{\mathrm{i}},\ i = 1, 2, \ldots, N \\ y \text{ is } B \end{array} \tag{1}$$

where B is a fuzzy set of conclusion to be determined. A and A_i are defined in a finite input space $\mathbf{X}$, $\dim(\mathbf{X}) = n$, while B_i and B are expressed in the output space $\mathbf{Y}$ with dimensionality, $\dim(\mathbf{Y}) = m$. The set of indexes of the rules is denoted by $\mathbf{N}$; in this case, it is simply a set of N natural numbers indexing the rules, $\mathbf{N} = \{1, 2, \ldots, N\}$.

There is a wealth of realizations of the inference schemes with a large number of optimization mechanisms, see (Oliveira et al. 2010; Apolloni et al. 2016; Alcala et al. 2009). In a nutshell, though the inference scheme is realized by determining the activation levels of the individual rules (their condition parts) implied by some A. This is typically done by computing a possibility measure of A and A_i, poss(A, A_i). Denoting the possibility value by λ_i, the conclusion B is taken as a union of B_i weighted by the activation levels (possibility values), namely

$$B(y) = \max_{i=1,2,\ldots,N} (\lambda_i(x) \wedge B_i(y)) \tag{2}$$

where $\wedge$ stands for the minimum operation. There are numerous variations of this inference scheme; nevertheless, the essence of the underlying reasoning process remains the same. Let us also stress that the result of inference is a fuzzy set.

Now, let us envision that instead of the entire collection of rules, we consider a subset of $\mathbf{I}$ rules in anticipation that this smaller collection can be deemed sufficient as being formed by a collection of the most representative rules out of N rules. Of course, the term representativeness has to be clarified and quantified as well as made operational. What is also quite intuitive is a fact that the rules forming the subset need to be made more abstract to compensate for the fact that they need to capture the entire set. Operationally, by making them more abstract (general) means that we form the condition parts of the selected rules more general. This, in effect, implies that instead of A_i occurring in the selected rule, we consider a certain granular abstraction of A_j, say $\boldsymbol{G}(A_j)$, where $\boldsymbol{G}(.)$ stands for the granular version of A_j. All in all, this generalization gives rise to the granular fuzzy rules

$$\text{if } G\left(A_{\mathrm{j}}\right) \text{ then } y \text{ } is \text{ } B_{\mathrm{j}} \tag{3}$$

$j = 1, 2, \ldots, I$. Now, $\mathbf{I}$ is a collection of "I" indexes coming from $\mathbf{N}$ identifying the subset of rules, that is $\mathbf{I} = \{j_1, j_2, \ldots, j_I\}$. The ensuing inference scheme comes in the form

$$\begin{array}{l} x \text{ is } A \\ \text{If } x \text{ is } \boldsymbol{G}(A_j) \text{ then } y \text{ is } B_j \\ y \text{ is } \boldsymbol{G}(B) y \text{ is } \boldsymbol{G}(B) \end{array} \tag{4}$$

It is worth noting that the granular format of the condition of the rule entails a granular format of the conclusion, so we obtain the granular counterpart $\boldsymbol{G}(B)$ instead of the fuzzy set B.

The granular version of A_j and $\boldsymbol{G}(A_i)$ can be articulated in different ways (Maciel et al. 2016; Crouzet and Strauss 2011; Zhang et al. 2011; Wilke and Portmann 2016). In a nutshell, the granularity of A_j results in non-numeric membership values. A granulation of membership function $\boldsymbol{G}(A_i)$ is a way of representing the unit interval of membership values as a finite and small collection of information granulation (Maciel et al. 2016). These information granules come with well-defined semantics, for example, *Low*, *Medium*, *High*, and *Very High* membership. A vocabulary comprising a finite number of information granules coming as a result of granulation $\boldsymbol{G}(A_i)$ is used as granulation representation of the original numeric membership grade. Several main alternatives are outlined in Table 1.

In the ensuing study, for the clarity of the presentation of the underlying concept and the overall methodology, our focus is on interval (set-based) granulation. Thus, we consider the interval-valued fuzzy sets, $\boldsymbol{G}(A_j)$ (see also Table 1). In this case in the general inference scheme (1), the activation of $\boldsymbol{G}(A_j)$ results in an interval of activation values $[\lambda_j^-, \lambda_j^+]$. As a result, the conclusion becomes an interval fuzzy set $[B_j^-, B_j^+]$ with the bounds computed as

$$[B_j^-(y)B_j^+(y)] = [\max_{j=1,2,\ldots,I}(\lambda_j^- \wedge B_j(y)), \max_{j=1,2,\ldots I}(\lambda_j^+ \wedge B_j(y)] \tag{5}$$

The development of the granular rule-based system entails two tightly connected design phases:

1. selection of the subset of rules **I** out of the entire collection of rules.
2. generalization of the condition parts of the rules [fuzzy sets A_j are made granular $\boldsymbol{G}(A_j)$].

These two steps are intertwined and have to be discussed together. The first one is evidently of structural (combinatorial) character. The second one is about making the original fuzzy sets of condition granular.

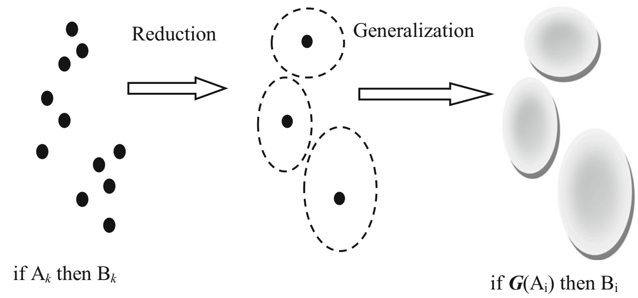

Fig. 1 Reduction of rule base by selection and a granular extension (generalization) of the representative subset of rules. The granular constructs are shown as *shadowed disks*

Figure 1 illustrates granular fuzzy rules in general. In the figure, we can visualize the process of the rule reduction by selecting subset of rules and the process of rule generalization by the constructing of the granular rules.

3 Designing granular rule-based system

The important issue in designing the granular rule-based system is how to construct the interval-valued membership function (Dubois and Prade 2016; Mendel 2015). The available information granularity (the level of granularity) is the most important asset and has to be carefully distributed among all the point in the membership function, so that the granular rules can covers (include) the unselected rules. In what follows, we propose several protocols of allocation of information granularity and discuss the indices, whose optimization is realized through this allocation.

The granular rules are formed by generalizing, via forming granular fuzzy sets in the condition parts of the rules. The process of forming $\boldsymbol{G}(A_i)$ out of A_i is realized through the allocation of information granularity. It is realized in several different ways. We discuss the performance of each of the protocol in the context of rules if x is $A_{i,}$ then y is B_i. Recall that the dimensionality of the input space is "n", while the output space has "m" elements. Several protocols of allocation of information granularity are outlined:

Table 1 Selected formal models of granular versions of fuzzy set A—membership grade $A(x)$ for fixed element of the universe of discourse

Interval granulation	$\boldsymbol{G}(A(x)) = [a_1(x), a_2(x)]$
Fuzzy set-based granulation	$\boldsymbol{G}(A(x)) = F_{A(x)}(u)$, $u \in [0,\ 1]$ where F is a fuzzy set defined in the unit interval
Probability-based granulation	$\boldsymbol{G}(A(x)) = \mathrm{p}_{A(x)}(u)$, $u \in [0,\ 1]$ where p is a probability density function defined in the unit interval, $\int_0^1 p_{A(x)}(u)\mathrm{d}u = 1$

Protocol 1 (P_1): a uniform allocation of information granularity for all membership degrees for the selected rules. The membership grades are replaced by an interval of the length α. More specifically, if a is the value of the membership grade, $a \in [0, 1]$, then the corresponding interval of membership values is expressed as $[a - \alpha/2, a + \alpha/2]$. We require that an overall balance of information granularity regarded as a modeling asset is satisfied, meaning that the sum of granularities is $n\alpha$. No optimization procedure is required.

Protocol 2 (P_2): a uniform allocation of information granularity with asymmetric position of interval. It is similar to P_1, however, it exhibits more flexibility, as we allow an asymmetric allocation of information granules (intervals) meaning that the membership values are now transformed to the interval $[a - \gamma\alpha, a + (1 - \gamma)\alpha]$, where $\gamma \in [0, 1]$. The optimization concerns an adjustment of the value of asymmetricity (γ). If $\gamma = 1/2$, the first protocol is a special case of this one.

Protocol 3 (P_3): it comes as an augmentation of P_2. We admit asymmetric allocation of information granularity to individual membership grades. The membership grades a_i, $i = 1, 2, \ldots, n$ are generalized and assuming the form of the interval $[a_i - \gamma_i\alpha, a_i + (1 - \gamma_i)\alpha]$, where $\gamma_i \in [0, 1]$. In total, we have a vector of coefficients $[\gamma_1, \gamma_2, \ldots, \gamma_n]$.

Protocol 4 (P_4): a non-uniform allocation of information granularity with symmetrically distributed intervals of information granules.

Here, the protocol involves individual intervals distributed symmetrically around a_i. They are formed as follows:

$$[a_i - \alpha_i/2, a_i + \alpha_i/2]. \tag{6}$$

The balance of information granularity is retained meaning that

$$\sum_{i=1}^{n} \alpha_i = n \cdot \alpha \tag{7}$$

Protocol 5 (P_5): a non-uniform allocation of information granularity with asymmetrically distributed intervals of information granules. Here, the protocol generalizes P_3 in the sense that the constructed intervals are distributed asymmetrically. Thus, α_i is replaced by the interval

$$[a_i - \alpha_i^-, a_i + \alpha_i^+] \tag{8}$$

with the balance of information granularity expressed as

$$\sum_{i=1}^{n} \alpha_i^- + \sum_{i=1}^{n} \alpha_i^+ = n \cdot \alpha. \tag{9}$$

In summary, the search space explored by each of the protocols can be described as follows.

Protocol	Parameters	Dimensionality of the search space
P_1	α	No optimization
P_2	γ, α	Optimization of γ, $\gamma \in [0, 1]$, (1)
P_3	α, γ_i $i = 1, 2, \ldots, n$	Optimization of $\gamma_1, \gamma_2, \ldots, \gamma_n$, ($n$)
P_4	α_i $i = 1, 2, \ldots, n$	Optimization of $\alpha_1, \alpha_2, \ldots \alpha_n$, ($n$)
P_5	α_i^-, α_i^+ $i = 1, 2, \ldots, n$	Optimization of $\alpha_1^-, \alpha_2^-, \ldots, \alpha_n^-$ and $\alpha_1^+, \alpha_2^+, \ldots, \alpha_n^+$, ($2n$)

Figure 2 shows the different between the original fuzzy rule-based system and the granular fuzzy rules-based system. The membership function for the original fuzzy rule-based system is depicted in Fig. 2a. Then, the granular fuzzy rules are achieved by shifting the points on the Gaussian function to the left and to the right based on the level of granularity, as shown in Fig. 2b.

When dealing with two-input (or multivariable) rule-based systems, the same protocols of allocation of information granularity apply, however, the condition on the retention of information granularity involves the sum $(n_1 + n_2)\alpha$, where n_1 and n_2 are the dimensionality of the corresponding input spaces say "if x is A_i and Z is C_i, then y is B_i". Here, A_i is defined over a discrete space dimensionality n_1, and C_i is expressed over a space of dimensionality n_2.

To complete optimization required by the protocols, we use the particle swarm optimization (PSO) techniques to search for the best subset of fuzzy rules and simultaneously realize the optimal allocation of information granulation,

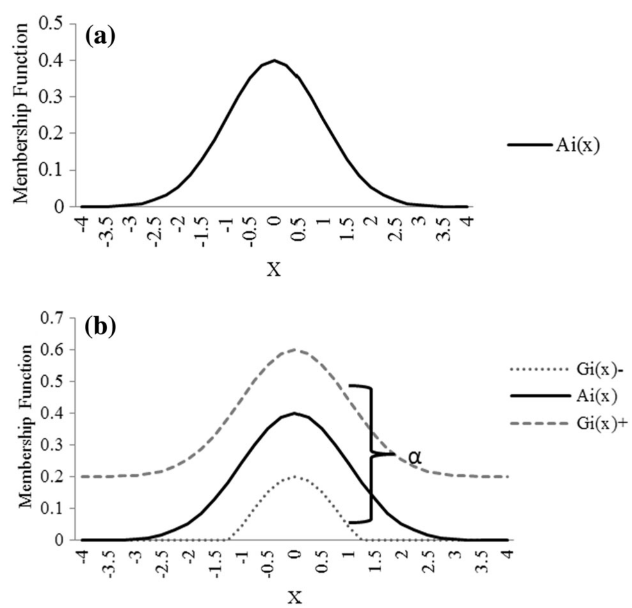

Fig. 2 **a** Example of an original fuzzy rule-based, $A_i(x)$, and **b** granular fuzzy rules $[G_i(x)^-, G_i(x)^+]$

so that the maximum the performance index (coverage) is satisfied.

4 Particle swarm optimization as a design environment

The optimization of granular fuzzy rules is completed in the setting of a certain information allocation protocol (Sakinah et al. 2013). For protocol P_1 and protocol P_2, we need to solve a single optimization task, namely, we have to select a subset of rules, $\mathbf{I} = \{j_1, j_2, \ldots, j_I\}$. Whereas, for protocols P_3, P_4, and P_5, we involve an additional optimization by completing an optimal allocation of information granularity. These two optimization tasks can be handled by the corresponding nested optimization process. In other words, for a selected subset of the rules generated by the optimization process involved at the upper level, we next carry out the optimal allocation of information granularity. In this study, we implement the generic particle swarm optimization for granularity allocation for protocol P_1 and P_2. For the protocols P_3, P_4, and P_5, we engage cooperative particle swarm optimization in which cases both phases of the optimization task are realized simultaneously.

4.1 Particle swarm optimization and its variants

Particle swarm optimization (PSO) (Kennedy and Eberhart 1995) was inspired by a collective behavior of birds or fish. PSO is a population-based method, where each individual, referred to as a particle, represent a candidate solution for an optimization problem. Each particle proceeds through the search space at a given velocity **v** that is dynamically modified according to the own experience of the particle and results in its local best (lb) performance. The particle is also affected by others particles experience, called global best (gb) (Eberhart and Shi 2001; Hu et al. 2004). The underlying expression for the update of the velocity in successive generations reads as follows:

$$\begin{aligned} \mathrm{v}_{i,j}(t+1) = \mathrm{w}\mathrm{v}_{i,j}(t) + c_1 r_{1,i}(t)[\mathrm{lb}_{i,j}(t) - x_{i,j}(t)] \\ + c_2 r_{2,i}(t)[\mathrm{gb}_j(t) - x_{i,j}(t)] \end{aligned} \quad (10)$$

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1) \quad (11)$$

where $i = 1, 2, \ldots, s$ (s the number of particles) and $j = 1, 2, \ldots, N + n$ (the search space is equal to the sum of the dimensions of the overall number of fuzzy rules and the dimensionality of the input space). The inertia weight (w) is confined to the range [0, 1]; its values can decrease over time. The cognitive factor c_1 and social factor c_2 determine the relative impact coming from the particle's own experience and the local best and global best. r_1 and r_2 are random numbers drawn from a uniform distribution defined over the unit interval that bring some component of randomness to the search process.

To deal with the large search spaces present in protocols P_3, P_4, and P_5, we employed another version of PSO, cooperative particle swarm optimization (CPSO). The motivation behind the use of CPSO, as advocated in (van den Bergh and Engelbrecht 2004), is to deal effectively with the high dimensionality of the search space, which becomes a serious concern when a large number of rules with its large dimensionality are involved. This curse of dimensionality is a significant impediment negatively impacting the effectiveness of the standard PSO. The essence of the cooperative version of PSO is essentially a parallel search for optimal subset of rules and its optimal allocation of information granulation values. The cooperative strategy is achieved by dividing the candidate solution vector into components, called sub-swarm, where each sub-swarm represents a small portion of the overall optimization processes. By doing this, we implement the concept of divide and conquer to solve the optimization problem, so that the process will become more efficient and fast.

The cooperative search realized between sub-swarms is achieved by sharing the information of the global best position (P_{GB}) across all sub-swarm. Here, the algorithm has the advantage of taking two steps forward, because the candidate solution comes from the best position for all sub-swarm except only for the current sub-swarm being evaluated. Therefore, the algorithm will not spend too much time optimizing the rules or allocating granularity that have little effect on the overall solution. The rate at which each swarm converges to the solution is significantly faster than the rate of convergence reported for the generic version of the PSO.

4.2 Fitness function

Let us assume that the set of rules **I** have been already formed (we discuss this development in the successive sections). The quality of these rules can be evaluated as follows. We consider the remaining $N–I$ rules not present in the collection of rules being retained. We treat successive $A_j s$ present there as the inputs to the inference process (4). The result becomes an information granule, $\boldsymbol{G}(B_j)$. Intuitively, the quality of the granular rule-based system depends how well the information granule $\boldsymbol{G}(B_j)$ "covers" the original B_j considering that the granular rules form only a subset of the original rule base. The fundamental with this regard is the notion of coverage of the information granule and its quantification. We introduce the following coverage index (measure)

$$\kappa = \frac{\sum_y \sum_j \mathrm{incl}[B_j(y), G(B_j(y))]}{(N-I)m} \quad (12)$$

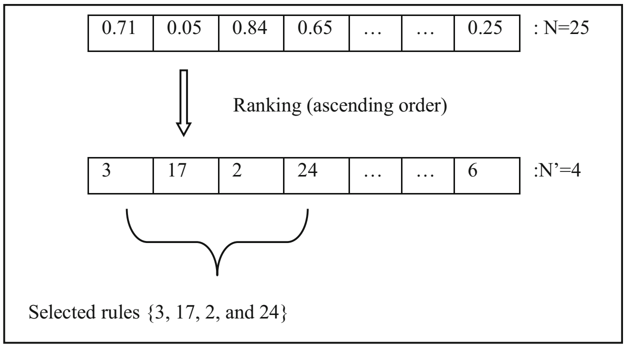

Fig. 3 From particle in the $[0, 1]^N$ search space to a subset of rules

where incl $[B_j(y), \boldsymbol{G}(B_j(y))]$ is a measure of inclusion of $B_j(y)$ in the granular counterpart produced by the inference scheme (4). The first summation standing in this formula is done over all elements of the finite output space over which B_j and $\boldsymbol{G}(B_j)$ are defined, whereas the second sum is carried out for all rules left out from the process of the generation of granular rules (whose number is $N–I$) and m is the dimension of the output space. The inclusion measure can be fully specified depending upon the assumed formalism used in the construction of granular rules. In the simplest case, where dealing with interval-valued membership functions, the double sum in the nominator of (12) is a count specifying how many times the membership grade $B_j(y)$ is contained in the interval.

Ideally, the coverage value is equal to 1, which becomes indicative of a complete inclusion of the conclusion (fuzzy set) of the original rule in the granular result of reasoning completed for the reduced rule base. In more realistic scenario, the ratio attains values lower than 1.

In addition, we introduce another objective function of a global character, called the area under the curve, AUC. As discussed above, the value of κ depends upon the predetermined level of granularity α, underlined here by the notation $\kappa(\alpha)$. The monotonicity property is apparent: $\kappa(\alpha)$ becomes a non-decreasing function of the level of granularity. To quantify an overall quality of the granular fuzzy rules, we integrate the corresponding values of $\kappa(\alpha)$, which results in a single index independent from the assumed level of granularity:

$$\mathrm{AUC} = \int_0^1 \kappa(\alpha)\mathrm{d}\alpha \tag{13}$$

Table 2 Collection of eight rules"if x is A_k, then y is B_k"

Rule	A_k	B_k
R_1	[0.1 0.9 0.5 0.2 0.1 0.0]	[0.0 0.3 0.5 0.8 1.0]
R_2	[0.7 1.0 0.6 0.3 0.2 0.0]	[1.0 0.7 0.3 0.2 0.0]
R_3	[0.9 0.9 1.0 0.2 0.0 0.0]	[0.1 0.9 0.9 0.4 0.2]
R_4	[0.0 0.3 0.5 0.9 1.0 0.0]	[0.0 0.4 0.9 1.0 0.5]
R_5	[1.0 0.9 0.5 0.2 0.1 0.0]	[0.0 0.3 0.5 0.8 1.0]
R_6	[0.6 0.3 0.2 1.0 0.5 0.7]	[0.5 0.9 1.0 0.5 0.2]
R_7	[0.2 0.3 1.0 0.2 0.5 0.7]	[0.0 0.3 0.5 0.8 1.0]
R_8	[0.0 1.0 0.5 0.3 0.0 0.0]	[0.3 1.0 0.2 0.0 0.0]

4.3 Particles representation

The first optimization phase is to select the optimal subset of rules. The problem is combinatorial in its nature. PSO is used here to form a subset of integers, which are the indexes of the rules to be used in the generation of the granular rules. As noted, **I** is represented as a set of indexes $\{j_1, j_2,\ldots, j_I\}$. The particle is then formed as a string of "N" real numbers positioned in [0, 1]. The search space is the hypercube $[0, 1]^N$. The particle is decoded as follows. The entries of the string are ranked. The result becomes a list of integers viewed as the indexes of the rules. The first I entries out of the N-position string are selected to form the subset of rules. Figure 3 illustrates the representation of the particle.

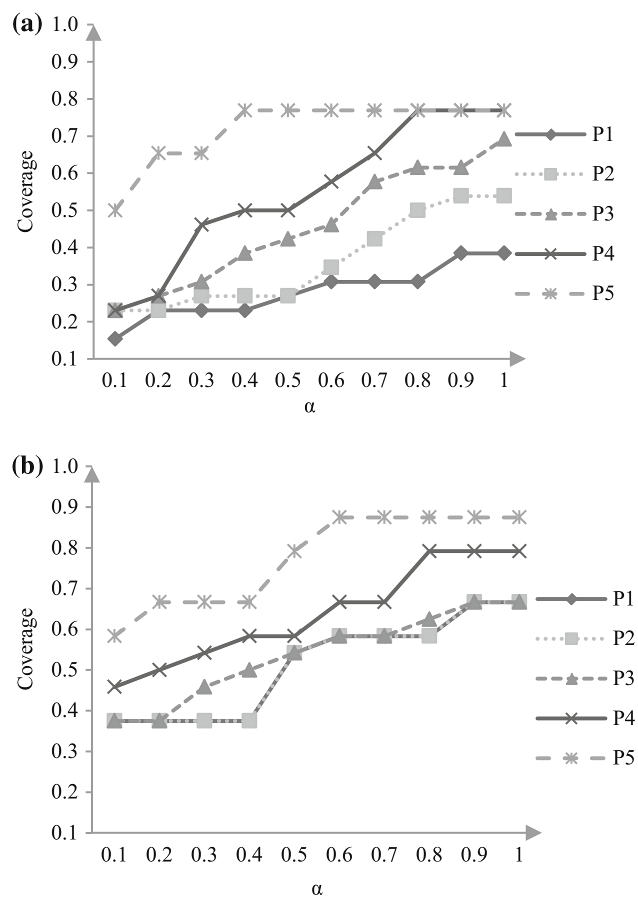

Fig. 4 Coverage produced by the five protocols, **a** two arbitrarily selected rules, and **b** optimized two rules

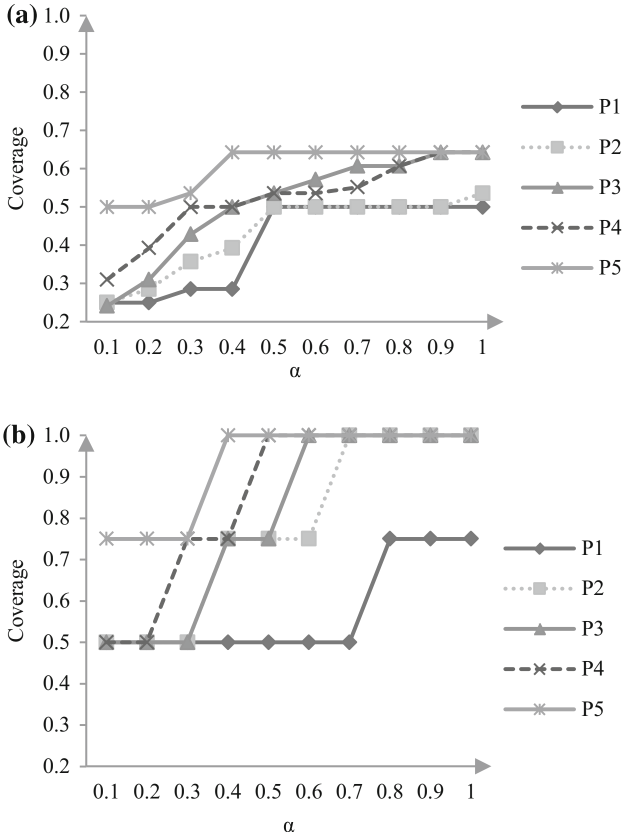

Fig. 5 Coverage produced by the five protocols, **a** $I = 1$ and **b** $I = 7$

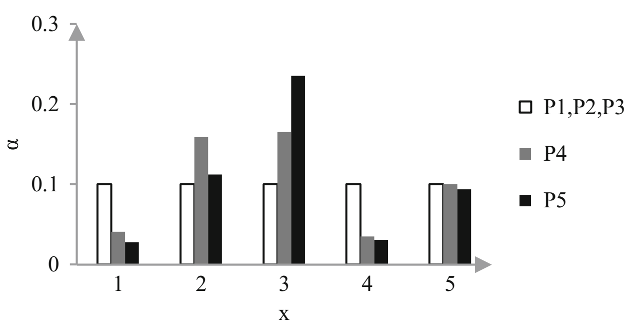

Fig. 6 Allocation of information granularity for $\alpha = 0.1$

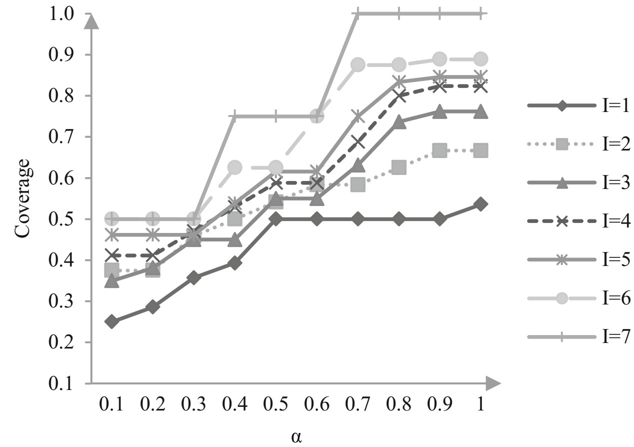

Fig. 7 Coverage versus different numbers of rules when using P_2

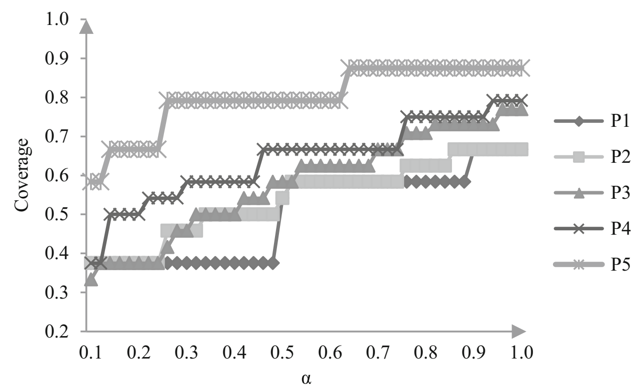

Fig. 8 Coverage versus α obtained for different protocols

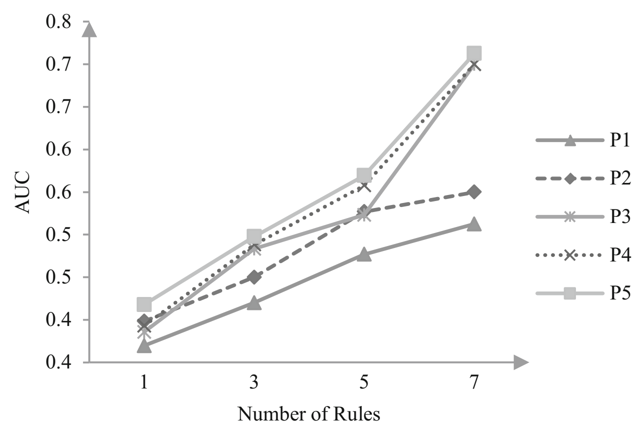

Fig. 9 AUC as a function of the number of rules

The second phase is to determine the optimal values of levels of information granulation. The optimization depends upon the protocol being used. In protocol P_3, PSO is used to find for best asymmetricity value, γ_i where $i = 1, 2, \ldots, n$ and $\gamma_i \in [0, 1]$. The particle is represented by a vector of numbers in [0, 1]. Its length is equal to the dimensionality of the finite universe of discourse over which the condition fuzzy sets are defined, namely, "n". The PSO used in the implementation of protocol P_4 similar to that used in the case of protocol P_3. The difference is the representation of each element is the allocation of the information granulation, α_i, where $i = 1, 2, \ldots, n$. Each element in the particle is represented by a real number that follows the constraint given by (8). Finally, in protocol P_5, the representation is almost the same as in protocol P_4. However, the length is two times higher as used in the previous protocol.

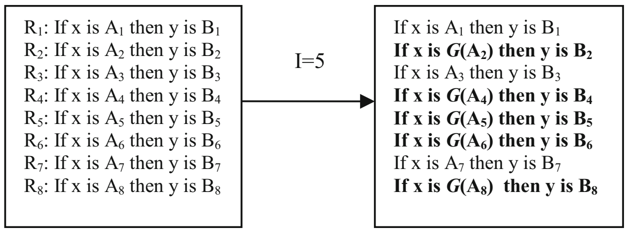

Fig. 10 Selected subsets of rules (in *boldface*) obtained for different numbers of selected rules (for protocol P_3)

Table 3 Rules for mortgage loan assessment

If (Asset is Low) and (Income is Low), then (Application is Low)
If (Asset is Low) and (Income is Medium), then (Application is Low)
If (Asset is Low) and (Income is High), then (Application is Medium)
If (Asset is Low) and (Income is Very High), then (Application is High)
If (Asset is Medium) and (Income is Low), then (Application is Low)
If (Asset is Medium) and (Income is Medium), then (Application is Medium)
If (Asset is Medium) and (Income is High), then (Application is High)
If (Asset is Medium) and (Income is Very High), then (Application is High)
If (Asset is High) and (Income is Low), then (Application is Medium)
If (Asset is High) and (Income is Medium), then (Application is Medium)
If (Asset is High) and (Income is High), then (Application is High)
If (Asset is High) and (Income is Very High), then (Application is High)

5 Experimental studies

In this section, we present a series of numeric experiments to illustrate the proposed method by showing its development, and quantifying the resulting performance. The experimental studies are concerned with the fuzzy rule-based systems applications. In all the experiment, we use the tenfold cross-validation method. We start the experiment by constructing the granular membership function by running the protocols of information granularity allocation, as presented in Sect. 3. The setup of the PSO is as follows: the number of generations is 200, and the size of particle is 100. The inertia weight, "w" changes linearly from 1 to 0 over the course of optimization. The values of the cognitive factor c_1 and social factor c_2 were set to 2.8 and 1.3, respectively (Eberhart and Shi 2001).

The performance of the granular rule-based system at the global level is based on the area under curve (AUC) for the coverage plot. The AUC is used to calculate the area of a region in the xy plane bounded by the graph of an

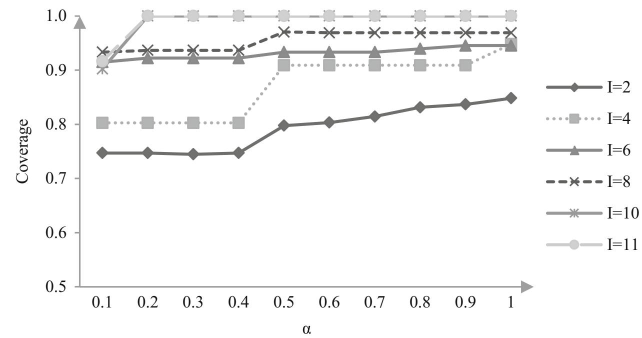

Fig. 11 Coverage produced by the different numbers of rules using P_5

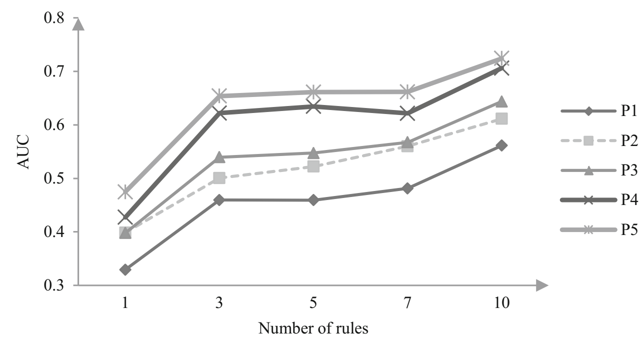

Fig. 12 Area under curve AUC

objective function called the coverage, the x-axis is the ε, and the vertical lines $\varepsilon = 0$ and $\varepsilon = 1$. As mentioned in the coverage value is in the interval [0, 1], therefore, the areas being all above the x-axis. The performance of the granular rule-based system is quantified by the values of AUC. Thus, we can investigate the performance of four methods at the global instead of local.

5.1 Synthetic fuzzy rule-based system

We consider the collection of eight rules (Table 2) "if x is A_k, then y is B_k" with fuzzy sets in the condition and conclusion part defined in the finite universes of discourse.

To illustrate the performance of the method, we start with a reduced set of two rules, $\mathbf{I} = \{7, 8\}$. These two rules were selected in an arbitrary fashion. The results are reported in Fig. 4a. There is a significant improvement when using protocol P_5 when compared the obtained results to the results produced by the remaining protocols. This is not surprising, as this protocol offers a significant level of flexibility when allocating information granularity. The improvement is particularly visible for low values of α.

Figure 4b shows the result using the optimal subset of two rules. Again, there is a visible improvement in comparison with the results presented in Fig. 4a.

Table 4 Rules for the aircraft landing control problem

1	If (Height is *L*) and (Velocity is DL), then (Control force is *Z*)
2	If (Height is *L*) and (Velocity is DS), then (Control force is DS)
3	If (Height is *L*) and (Velocity is *Z*), then (Control force is DL)
4	If (Height is *L*) and (Velocity is US), then (Control force is DL)
5	If (Height is *L*) and (Velocity is UL), then (Control force is DL)
6	If (Height is *M*) and (Velocity is DL), then (Control force is US)
7	If (Height is *M*) and (Velocity is DS), then (Control force is *Z*)
8	If (Height is *M*) and (Velocity is *Z*), then (Control force is DS)
9	If (Height is *M*) and (Velocity is US), then (Control force is DL)
10	If (Height is *M*) and (Velocity is UL), then (Control force is DL)
11	If (Height is *S*) and (Velocity is DL), then (Control force is UL)
12	If (Height is *S*) and (Velocity is DS), then (Control force is US)
13	If (Height is *S*) and (Velocity is *Z*), then (Control force is *Z*)
14	If (Height is *S*) and (Velocity is US), then (Control force is DS)
15	If (Height is *S*) and (Velocity is UL), then (Control force is DL)
16	If (Height is NZ) and (Velocity is DL), then (Control force is UL)
17	If (Height is NZ) and (Velocity is DS), then (Control force is UL)
18	If (Height is NZ) and (Velocity is *Z*), then (Control force is *Z*)
19	If (Height is NZ) and (Velocity is US), then (Control force is DS)
20	If (Height is NZ) and (Velocity is UL), then (Control force is DS)

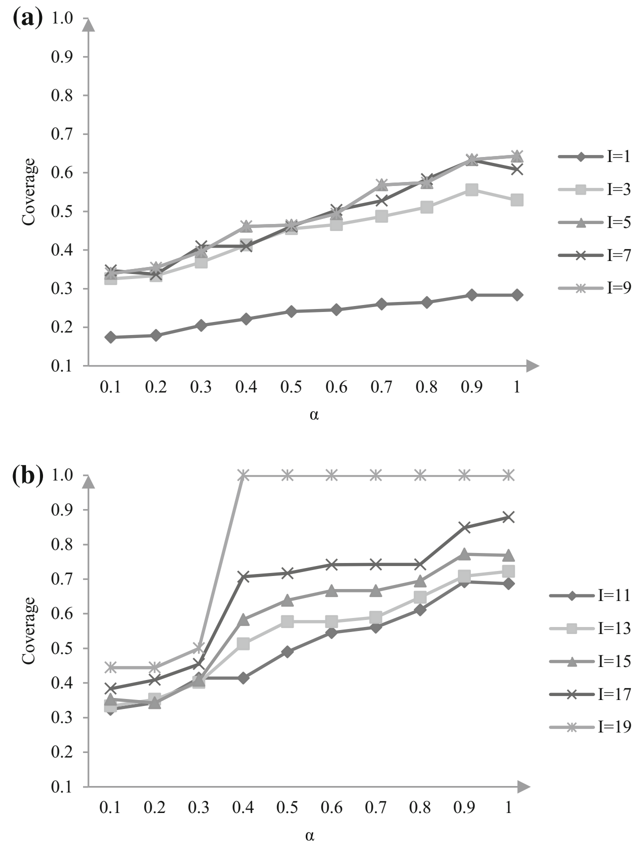

Fig. 13 Plot of coverage $\kappa(\alpha)$ regarded as a function of using P_4

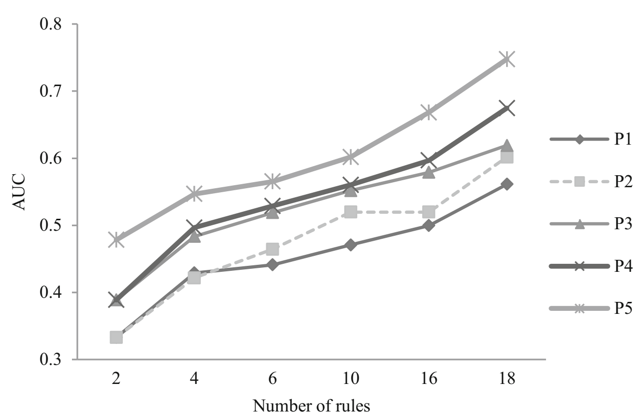

Fig. 14 Area under curve AUC

Figure 5 illustrates the coverage values when using the PSO-optimized subsets of rules with $I = 1$ and 7. The quality of results (ranging from the weakest coverage to the highest one) brings a ranking of the protocols ordered as P_1, P_2, P_3, P_4, and P_5 with P_1 producing the lowest coverage.

In the sequel, Fig. 6 shows a distribution of the allocation of granularity realized with the use of the protocol P_5; apparently, the distribution becomes non-uniform over the input space.

Figure 7 illustrates the values of coverage when using different numbers of rules. The coverage values are higher when increasing the number of selected rules. As illustrated in Fig. 8, protocols of higher flexibility produce better coverage results.

The overall performance expressed in terms of the AUC values is visualized in Fig. 9. Again, the superiority of the most flexible protocols is visible. The reduced list of rules is presented in Fig. 10.

5.2 Mortgage applications assessment rule-based system

Assessment of a mortgage application normally based on evaluating the market value and location of the house, the applicant's asset and income, and repayment plan. A collection of rules is shown in Table 3.

The results expressed in terms of the coverage treated as a function of the number of retained rules are summarized in Figs. 11 and 12. The main trends are apparent. Furthermore, the quantification of the improvements resulting from the increase of the number of rules involved is visible; a substantial jump is present when using more than four rules.

Table 5 Rules for the service center

1	If (Mean_Delay is VS) and (# of server is *S*) and (Utilization_Factor is *L*), then (# of spare is VS)
2	If (Mean_Delay is *S*) and (# of server is *S*) and (Utilization_Factor is *L*), then (# of spare is VS)
3	If (Mean_Delay is *M*) and (# of server is *S*) and (Utilization_Factor is *L*), then (# of spare is VS)
4	If (Mean_Delay is VS) and (# of server is *M*) and (Utilization_Factor is *L*) then (# of spare is VS)
5	If (Mean_Delay is *S*) and (# of server is *M*) and (Utilization_Factor is *L*), then (# of spare is VS)
6	If (Mean_Delay is *M*) and (# of server is *M*) and (Utilization_Factor is *L*), then (# of spare is VS)
7	If (Mean_Delay is VS) and (# of server is *L*) and (Utilization_Factor is *L*), then (# of spare is *S*)
8	If (Mean_Delay is *S*) and (# of server is *L*) and (Utilization_Factor is *L*), then (# of spare is *S*)
9	If (Mean_Delay is *M*) and (# of server is *L*) and (Utilization_Factor is *L*), then (# of spare is VS)
10	If (Mean_Delay is VS) and (# of server is *S*) and (Utilization_Factor is *M*), then (# of spare is *S*)
11	If (Mean_Delay is *S*) and (# of server is *S*) and (Utilization_Factor is *M*), then (# of spare is *S*)
12	If (Mean_Delay is *M*) and (# of server is *S*) and (Utilization_Factor is *M*), then (# of spare is VS)
13	If (Mean_Delay is VS) and (# of server is *M*) and (Utilization_Factor is *M*), then (# of spare is RS)
14	If (Mean_Delay is *S*) and (# of server is *M*) and (Utilization_Factor is *M*), then (# of spare is *S*)
15	If (Mean_Delay is *M*) and (# of server is *M*) and (Utilization_Factor is *M*), then (# of spare is VS)
16	If (Mean_Delay is VS) and (# of server is *L*) and (Utilization_Factor is *M*), then (# of spare is M)
17	If (Mean_Delay is *S*) and (# of server is *L*) and (Utilization_Factor is *M*), then (# of spare is RS)
18	If (Mean_Delay is *M*) and (# of server is L) and (Utilization_Factor is *M*), then (# of spare is *S*)
19	If (Mean_Delay is VS) and (# of server is *S*) and (Utilization_Factor is *H*), then (# of spare is VL)
20	If (Mean_Delay is *S*) and (# of server is *S*) and (Utilization_Factor is *H*), then (# of spare is *L*)
21	If (Mean_Delay is *M*) and (# of server is *S*) and (Utilization_Factor is *H*), then (# of spare is *M*)
22	If (Mean_Delay is VS) and (# of server is *M*) and (Utilization_Factor is *H*), then (# of spare is *M*)
23	If (Mean_Delay is *S*) and (# of server is *M*) and (Utilization_Factor is *H*), then (# of spare is *M*)
24	If (Mean_Delay is *M*) and (# of server is *M*) and (Utilization_Factor is *H*), then (# of spare is *S*)
25	If (Mean_Delay is VS) and (# of server is *L*) and (Utilization_Factor is *H*), then (# of spare is RL)
26	If (Mean_Delay is *S*) and (# of server is *L*) and (Utilization_Factor is *H*), then (# of spare is *M*)
27	If (Mean_Delay is *M*) and (# of server is *L*) and (Utilization_Factor is *H*), then (# of spare is RS)

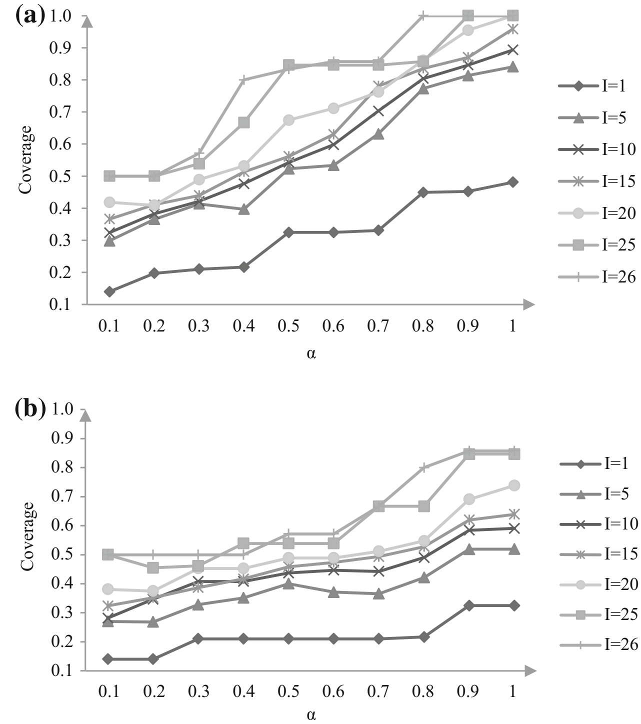

Fig. 15 Plot of coverage $\kappa(\alpha)$ regarded as a function of α: **a** P_1 and **b** P_2

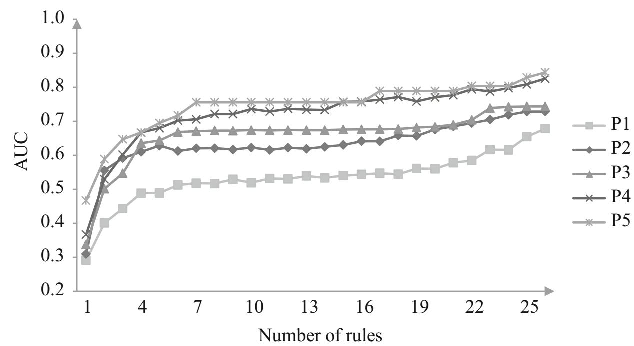

Fig. 16 Area under curve versus the number of rules

5.3 Aircraft landing control problem

The aircraft landing control problem is dealing with the two important parameters called the velocity and the height. The main objective is to control the landing approach of an aircraft by desired downward velocity that is proportional to the square of the height. For example, at higher altitudes, a large downward velocity is desired, and as the altitude (height) diminishes, the desired downward velocity gets smaller and smaller. Finally, as the height becomes vanishingly small, the downward velocity also goes to zero. Therefore, the aircraft will descend form

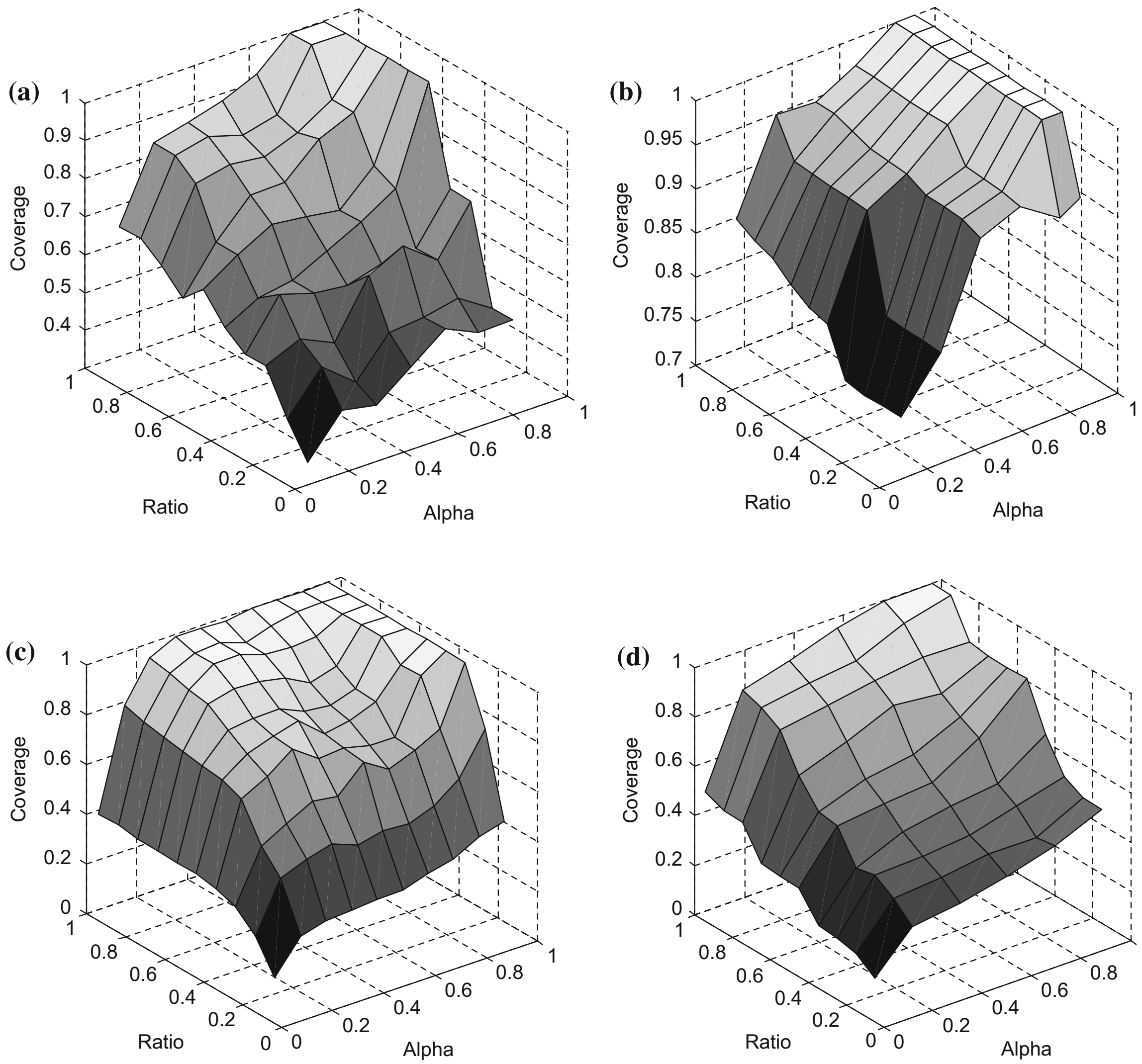

Fig. 17 Coverage as a function of the fraction of rules retained for data: **a** synthetic, **b** applicant, **c** aircraft, and **d** service. In all the cases, protocol P_5 was used

altitude promptly, so that the touch down process is very gently to avoid damage. The pertinent rules are shown in Table 4.

The main results are summarized in Figs. 13 and 14.

5.4 Service center operation data

The rules having three inputs and a single output describing the functioning of the center are presented in Table 5. The overall number of the rules is 27.

The summary of the results is presented in Figs. 15 and 16.

A concise summary of the results obtained for the series of experiments is presented in Fig. 17. Here, we visualize the coverage as a function of a fraction of rules retained (ratio). While the monotonicity character of this relationship is visible, these plots show how the changes are distributed.

6 Conclusions

The general issue of structural compression of rule-based systems was presented as inherently associated with the emergence of granular constructs. Information granularity is reflective of the increased level of abstraction of the reduced set of rules. Information granularity is sought as an essential asset, whose prudent allocation is behind the design of optimally reduced rule-based systems. The experimental part of the study shows essential linkages among the quality of the granular fuzzy rules and the number of retained rules and the admitted level of information granularity.

It has to be noted that the granular fuzzy sets form a general concept; however, in this study, we focused on their interval realization. The entire development was presented in this way for clarity purposes (given our intent to concentrate on the concept). Nevertheless, considerations of other realizations of the granular constructs follow the same general scheme and require some slight modifications.

Acknowledgments This research was supported by Universiti Teknikal Malaysia Melaka and Ministry of Education Malaysia under Grant No. FRGS/2013/FTMK/ICT02/02/1/F00161.

References

Alcala R et al (2009) A multiobjective evolutionary approach to concurrently learn rule and data bases of linguistic fuzzy-rule-based systems. IEEE Trans Fuzzy Syst 17(5):1106–1122

Antonelli M et al (2016) Multi-objective evolutionary design of granular rule-based classifiers. Granul Comput 1(1):37–58. doi:10.1007/s41066-015-0004-z

Apolloni B et al (2016) A neurofuzzy algorithm for learning from complex granules. Granul Comput. doi:10.1007/s41066-016-0018-1

Baranyi P, Yam Y. (2000) Fuzzy rule base reduction. In fuzzy IF-THEN rules in computational intelligence: theory and applications. pp 135–160. http://www.scopus.com/inward/record.url?eid=2-s2.0-0002973811&partnerID=tZOtx3y1

Cordon O (2011) A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: designing interpretable genetic fuzzy systems. Int J Approx Reason 52(6):894–913

Crouzet JF, Strauss O (2011) Interval-valued probability density estimation based on quasi-continuous histograms: proof of the conjecture. Fuzzy Sets Syst 183(1):92–100

Dubois D, Prade H (2016) Bridging gaps between several forms of granular computing. Granul Comput 1(2):115–126. doi:10.1007/s41066-015-0008-8

Eberhart RC, Shi Y (2001). Particle swarm optimization: developments, applications and resources. In: Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546), 1, pp 81–86. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=934374

Gacto MJ, Alcalá R, Herrera F (2011) Interpretability of linguistic fuzzy rule-based systems: an overview of interpretability measures. Inf Sci 181(20):4340–4360. http://www.sciencedirect.com/science/article/pii/S0020025511001034. Accessed February 9, 2015

Hu X, Shi Y, Eberhart R (2004) Recent advances in particle swarm. IEEE Congr Evolut Comput 1:90–97

Juang C-F, Chen C-Y (2013) Data-driven interval type-2 neural fuzzy system with high learning accuracy and improved model interpretability. IEEE Trans Cybern 43(6):1781–95. http://www.ncbi.nlm.nih.gov/pubmed/24273147

Kennedy J, Eberhart R (1995) Particle swarm optimization. Proc IEEE Int Conf Neural Netw 4:1942–1948

Maciel L, Ballini R, Gomide F (2016) Evolving granular analytics for interval time series forecasting. Granul Comput. doi:10.1007/s41066-016-0016-3

Mendel JM (2015) A comparison of three approaches for estimating (synthesizing) an interval type-2 fuzzy set model of a linguistic term for computing with words. Granul Comput 1(1):59–69. doi:10.1007/s41066-015-0009-7

Oliveira DN, De Lima Henn GA, Da Mota Almeida O (2010) Design and implementation of a Mamdani fuzzy inference system on an FPGA using VHDL. In: Annual Conference of the North American Fuzzy Information Processing Society—NAFIPS, pp 1–6

Riid A, Rüstern E (2011) Identification of transparent, compact, accurate and reliable linguistic fuzzy models. Inf Sci 181(20):4378–4393. http://www.sciencedirect.com/science/article/pii/S0020025511000685. Accessed February 9, 2015

Sakinah S, Ahmad S, Pedrycz W (2013) Fuzzy rule-based system through granular computing. In Systems, Man, and Cybernetics (SMC), IEEE, pp 800–805. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6721894. Accessed December 31, 2014

van den Bergh F, Engelbrecht AP (2004) A cooperative approach to particle swarm optimization. IEEE Trans Evolut Comput 8(3):225–239. http://www.scopus.com/inward/record.url?eid=2-s2.0-3142697802&partnerID=tZOtx3y1. Accessed April 1, 2014

Villar P et al. (2012) Feature selection and granularity learning in genetic fuzzy rule-based classification systems for highly imbalanced data-sets. Int J Uncertain Fuzziness Knowl Syst 20(03):369–397. http://www.worldscientific.com/doi/abs/10.1142/S0218488512500195

Wilke G, Portmann E (2016) Granular computing as a basis of human–data interaction: a cognitive cities use case. Granul Comput. doi:10.1007/s41066-016-0015-4

Zhang M, Dong K, Yu F (2011) "Fuzzy granulation of interval numbers," Fuzzy Systems and Knowledge Discovery (FSKD), Eighth International Conference on, Shanghai, 2011, pp 372–376. doi:10.1109/FSKD.2011.6019540

Zhou S-M, Gan JQ (2008) Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling. Fuzzy Sets Syst 159(23):3091–3131. http://linkinghub.elsevier.com/retrieve/pii/S0165011408002765. Accessed October 22, 2013