Paper:

Independent Evaluations of Each Fuzzy Rule for Derivative-Free Optimization of Fuzzy Systems: Toward Fast Fuzzy-Rules Learning for Fuzzy Inputs

Kiyohiko Uehara* **and Kaoru Hirota****

* Ibaraki University, Hitachi 316-8511, Japan
E-mail: kiyohiko.uehara.art@vc.ibaraki.ac.jp
** Beijing Institute of Technology, Beijing 100081, China
[Received July 24, 2020; accepted August 15, 2020]

A method is proposed for evaluating fuzzy rules independently of each other in their optimization. It is derived by extending the conventional method called α-FUZZI-ES so as to cope with facts (inputs) given by fuzzy sets (non-singletons). A method is further proposed for fuzzy rules learning based on the evaluation method. It attains fast fuzzy-rules learning by optimizing fuzzy rules independently of each other in parallel. The proposed method is effective especially when evaluation functions for fuzzy rules learning are not differentiable and then derivative-free optimization is required. Numerical results indicate that the learning method achieves proper convergence with derivative-free optimization.

Keywords: fuzzy inference, fuzzy rules learning, fast optimization, parallel processing, derivative-free optimization

1. Introduction

Parallel fuzzy inference provides an effective scheme to represent complex nonlinear systems with a number of fuzzy rules. The fuzzy rules have been often optimized by learning algorithms. Especially when evaluation functions for fuzzy rules learning are not differentiable, derivative-free methods are adopted, including genetic algorithms, particle swarm optimization, and artificial immune systems. Conventional methods for fuzzy rules learning have applied such optimization methods by using functions for evaluating overall performance of fuzzy systems. As the number of fuzzy rules increases, the search space of their optimal parameters becomes larger. It entails the deceleration of convergence speed in fuzzy rules learning.

In order to solve the problem mentioned above, α-FUZZI-ES (α-weight-based fuzzy-rule independent evaluations) has been proposed for evaluating fuzzy rules independently of each other [1]. Moreover, α-FUZZI-ES learning (α-FUZZI-ES-based fuzzy-rule learning) has been proposed for fuzzy rules learning by the effective use of α-FUZZI-ES [1]. It can optimize fuzzy rules independently of each other in parallel. Thereby, it attains fast fuzzy-rules learning with parallel processing. α-FUZZI-ES learning also reduces the dimensionality of the solution space for finding the optimal fuzzy rules and thus it makes the convergence speed fast in fuzzy rules learning. α-FUZZI-ES learning is effective for tuning each fuzzy rule with derivative-free optimization especially when the evaluation functions are not differentiable and then derivative-based optimization methods cannot be applied. It provides a way to tune fuzzy rules independently of each other even with derivative-free optimization. α-FUZZI-ES learning, however, requires the condition that facts (inputs) are given by singletons and antecedent fuzzy sets form a strong fuzzy partition.

This paper proposes a method to solve the above-mentioned problem by extending α-FUZZI-ES so as to cope with facts given by fuzzy sets in the meaning of non-singletons. The proposed method is named α-FUZZI-EX (α-weight-based fuzzy-rule independent evaluations extended for fuzzy inputs) in this paper. A method is further proposed for attaining fast fuzzy-rules learning by applying α-FUZZI-EX. It is named α-FUZZI-EX learning (α-FUZZI-EX-based fuzzy-rule learning) in this paper. Numerical results demonstrate the proper convergence in fuzzy rules learning with derivative-free optimization and facts given by fuzzy sets (non-singletons).

2. Definitions and Preliminaries

For the following discussions, some definitions and preliminaries are presented. Further details of each are described in [2–4].

Definition 1 When a convex fuzzy set A in the universe of discourse X is defined by a continuous membership function $\mu_A(x)\ (x \in X)$ and its α-cuts (also called α-level sets) are all bounded, *the reference point* x_A° of A is defined by using its α-cut A_α as follows:

$$x_A^\circ = \frac{x_\alpha^\ell + x_\alpha^u}{2}, \quad \alpha = \max_x \mu_A(x), \qquad (1)$$

where x_α^u and x_α^ℓ denote the least upper and the greatest lower bounds of A_α, respectively. ■

Definition 2 Suppose that a convex fuzzy set A in the universe of discourse X is defined by a continuous membership function $\mu_A(x)\ (x \in X)$ and its α-cuts are all bounded. The fuzzy set A is *symmetric* if and only if the following equation holds:

$$\frac{x_\alpha^\ell + x_\alpha^u}{2} = x_A^\circ, \quad \forall \alpha \in (0, \alpha_{\max}], \ \alpha_{\max} = \max_x \mu_A(x). \quad (2)$$

Here, x_α^u and x_α^ℓ denote the least upper and the greatest lower bounds of the α-cut A_α of A, respectively. The symbol x_A° represents the reference point of A. A convex fuzzy set is called *asymmetric* if and only if Eq. (2) does not hold. ■

Definition 3 The generalized mean $M(\{x_j, p_j\}; \omega)$ is defined by

$$M(\{x_j, p_j\}; \omega) = \left[\frac{\sum_{j=1}^{n} p_j x_j^\omega}{\sum_{j=1}^{n} p_j} \right]^{\frac{1}{\omega}}, \quad x_j > 0, \quad p_j > 0, \quad (3)$$

where x_j denotes a real number in the universe of discourse and p_j represents a real number used for the weight of x_j. The symbol ω denotes a real number to determine the property of the mean [2, 3]. ■

3. Independent Evaluations of Each Fuzzy Rule and Their Use for Fuzzy Rules Learning

This section first introduces parallel fuzzy inference for the following discussions. Then, methods are proposed for evaluating fuzzy rules independently of each other and for fuzzy rules learning.

3.1. Parallel Fuzzy Inference

This paper treats the parallel fuzzy inference in the form below:

$$\begin{array}{rl} \text{Rule 1:} & \text{If } x \text{ is } P_1 \text{ then } y \text{ is } Q_1. \\ \text{Rule 2:} & \text{If } x \text{ is } P_2 \text{ then } y \text{ is } Q_2. \\ \vdots & \qquad \vdots \\ \text{Rule } n\text{:} & \text{If } x \text{ is } P_n \text{ then } y \text{ is } Q_n. \\ \text{Given fact:} & x \text{ is } \tilde{P}. \\ \hline \text{Consequence:} & y \text{ is } \tilde{Q}. \end{array}$$

Here, P_j and $\tilde{P}$ denote fuzzy sets in the universe of discourse X, whereas Q_j and $\tilde{Q}$ represent fuzzy sets in the universe of discourse Y. In particular, P_j in the antecedent part of the fuzzy rule is called *an antecedent fuzzy set*, whereas Q_j in the consequent part of the fuzzy rule is called *a consequent fuzzy set*. In this paper, P_j, Q_j, and $\tilde{P}$ are all defined by normal and convex fuzzy sets and their reference points are placed in $[0, 1]$. The membership functions of P_j, Q_j, $\tilde{P}$, and $\tilde{Q}$ are respectively denoted by $\mu_{P_j}(x)$, $\mu_{Q_j}(y)$, $\mu_{\tilde{P}}(x)$, and $\mu_{\tilde{Q}}(y)$, where $x \in X$ and $y \in Y$. For convenience in the following discussions, the j-th fuzzy rule is represented by R_j.

3.2. Mathematical Derivation of a Method for Independent Evaluations of Each Fuzzy Rule

In this section, a method is proposed for evaluating fuzzy rules independently of each other in fuzzy rules learning, especially with facts given by fuzzy sets (non-singletons). It is derived in a way that extends α-FUZZI-ES. α-FUZZI-ES is effectual only under condition that facts are given by singletons and antecedent fuzzy sets form a strong fuzzy partition. The validity of the proposed method is mathematically proven.

For the following discussions, suppose that the conditions below are satisfied:

(1a) Learning data are given by the input–output pairs $(\hat{P}_k, \hat{Q}_k)$ $(k = 1, 2, \ldots, n_d)$, where $\hat{P}_k$ and $\hat{Q}_k$ respectively denote the input and output fuzzy sets of the learning data.

(2a) The value E_k of an evaluation function for optimizing fuzzy rules with $(\hat{P}_k, \hat{Q}_k)$ satisfies the condition that $E_k \geq 0$. In this paper, the smaller value of E_k indicates higher performance in fuzzy rules learning.

(3a) The compatibility degree $\tilde{p}_{jk}$ between P_j and $\hat{P}_k$ is normalized as shown below:

$$\check{p}_{jk} = \frac{\tilde{p}_{jk}}{\sum_{j=1}^{n} \tilde{p}_{jk}}. \quad (4)$$

In particular, $\tilde{p}_{jk}$ is defined by

$$\tilde{p}_{jk} = \sup_x [\mu_{\hat{P}_k}(x) \wedge \mu_{P_j}(x)], \quad (5)$$

where $\mu_{\hat{P}_k}(x)$ denotes the membership function of $\hat{P}_k$ and $\wedge$ represents the minimum operation.

Note that P_j and $\tilde{P}$ can be multi-dimensional fuzzy sets in the following discussions.

From Eq. (4), the equation below holds obviously:

$$\sum_{j=1}^{n} \check{p}_{jk} = 1. \quad (6)$$

The value E of the evaluation function for all the learning data in fuzzy rules optimization can be formulated by using Eq. (6) as follows:

$$E = \sum_{k=1}^{n_d} E_k = \sum_{k=1}^{n_d} \left[E_k \sum_{j=1}^{n} \check{p}_{jk} \right] = \sum_{j=1}^{n} \left[\sum_{k=1}^{n_d} E_k \check{p}_{jk} \right] = \sum_{j=1}^{n} e_j, \quad (7)$$

where

$$e_j = \sum_{k=1}^{n_d} E_k \check{p}_{jk}. \quad (8)$$

Let L_j be defined by

$$L_j = \{(\hat{P}_k, \hat{Q}_k) \mid \tilde{p}_{jk} > 0, \ k = 1, 2, \ldots, n_d\}. \quad (9)$$

For the following discussions, L_j is represented by

$$L_j = \{(\hat{P}_{jk'}, \hat{Q}_{jk'})\}, \ k' = 1, 2, \ldots, n_{d,j}, \quad (10)$$

where $n_{d,j}$ represents the number of the learning data in L_j. Because $\check{p}_{jk} = 0$, $k \in \{k'' \mid (\hat{P}_{k''}, \hat{Q}_{k''}) \notin L_j\}$, Eq. (8) turns to

$$e_j = \sum_{k'=1}^{n_{d,j}} E_{jk'} \check{p}_{jk'}, \quad (11)$$

where $E_{jk'}$ denotes the value $E_{k'}$ of the evaluation function used in optimizing R_j with the learning data $(\hat{P}_{jk'}, \hat{Q}_{jk'})$.

In Eq. (11), $E_{jk'} \check{p}_{jk'}$ is in the form of $E_{jk'}$ weighted by $\check{p}_{jk'}$. Such weighting on the basis of compatibility degrees is called *α-weighting* [1]. The value of $\check{p}_{jk'}$ indicates the degree to which $\hat{P}_{jk'}$ belongs to P_j. From these viewpoints, it can be considered that $\check{p}_{jk'}$ divides out the value of $E_{jk'}$ among the adjacent fuzzy rules of R_j. Therefore, the value of e_j can be regarded as the evaluation value of R_j. The values of e_j $(j=1,2,\ldots,n)$ can be obtained independently of each other. This property makes it possible to calculate the values of e_j $(j=1,2,\ldots,n)$ in parallel and leads to fast computing in fuzzy rule evaluations.

The above-mentioned method for fuzzy rule evaluations is named *α-FUZZI-EX* (α-weight-based fuzzy-rule independent evaluations extended for fuzzy inputs) in this paper. α-FUZZI-ES is a special case of α-FUZZI-EX. α-FUZZI-EX is equivalent to α-FUZZI-ES under the following conditions [1]:

(1b) The antecedent fuzzy sets P_j $(j=1,2,\ldots,n)$ form a strong fuzzy partition. Namely, the following equation is satisfied:

$$\sum_{j=1}^{n}\mu_{P_j}(x)=1. \qquad (12)$$

(2b) The inputs of learning data are given by singletons.

Unlike α-FUZZI-ES, α-FUZZI-EX allows to cope with facts given by fuzzy sets (non-singletons) and does not require the antecedent fuzzy sets to form a strong fuzzy partition of X.

3.3. Fuzzy Rules Learning Based on α-FUZZI-EX

In this section, a fast method is proposed for fuzzy rules learning by the effective use of α-FUZZI-EX. The following equation holds because $E_{jk'}\geq 0$ and $\check{p}_{jk'}\geq 0$:

$$e_j\geq 0,\ \ j=1,2,\ldots,n. \qquad (13)$$

As can be found from Eqs. (7) and (13), E is obtained by the summation of the positive or zero values given by e_j. Accordingly, E can be minimized by decreasing each value of e_j $(j=1,2,\ldots,n)$. As discussed in Section 3.2, the values of e_j $(j=1,2,\ldots,n)$ can be calculated independently of each other. Therefore, fuzzy rules can be tuned independently of each other in parallel for their global optimization by minimizing the each value of e_j $(j=1,2,\ldots,n)$. α-FUZZI-EX attains fast fuzzy-rules learning with hardware in parallel including GPGPUs (general-purpose graphics processing units) and many-core CPUs (central processing units).

In order to evaluate R_j, the value of e_j is calculated by using the number $n_{d,j}$ of the learning data in L_j as shown in Eq. (11). When each fuzzy rule is required to be evaluated per one learning data, the following definition of fuzzy rule evaluation can be used on behalf of e_j:

$$\bar{e}_j=\frac{e_j}{n_{d,j}}. \qquad (14)$$

The above-mentioned fuzzy-rules learning on the basis of α-FUZZI-EX is named *α-FUZZI-EX learning* (α-FUZZI-EX-based fuzzy-rule learning) in this paper. α-FUZZI-EX learning inherits the practically important properties of α-FUZZI-ES learning including above-mentioned fast computing with hardware in parallel. The other properties are as follows: α-FUZZI-EX learning provides a self-governing scheme of each fuzzy rule in its optimization. It is effective especially when evolutionary algorithms are adopted for optimizing fuzzy rules independently of each other. α-FUZZI-EX learning can reduce the dimensionality of the solution space for finding the optimal fuzzy rules. These inherited properties are detailed in [1].

In the case where learning data with singleton inputs are given and antecedent fuzzy sets form strong fuzzy partition of X, α-FUZZI-EX learning provides exactly the same performance as α-FUZZI-ES learning. This is because α-FUZZI-ES is a special case of α-FUZZI-EX as described in Section 3.2 and then α-FUZZI-EX learning is equivalent to α-FUZZI-ES learning in the case.

4. Application of α-FUZZI-EX learning to Interval Prediction

α-FUZZI-EX learning is applied to interval prediction for demonstrating its performance. α-GEMII (α-level-set and generalized-mean-based inference with the proof of two-sided symmetry of consequences) [2–4] is adopted because it is effective to represent prediction intervals by its deduced consequences.

4.1. α-GEMII Specialized for Triangular Membership Functions

α-GEMII is applied to interval prediction, wherein Q_j $(j=1,2,\ldots,n)$ are optimized by α-FUZZI-EX learning. In the use of α-GEMII, antecedent and consequent fuzzy sets are set to be symmetric and are defined by triangular membership functions for convenience in interval prediction. The following specializes the operations in α-GEMII for triangular membership functions. The operational steps in the general form of α-GEMII are described in [2, 3].

Under the condition that P_j, Q_j, and $\tilde{P}$ are normal and are defined by triangular membership functions, α-GEMII can deduce consequences in the form of normal fuzzy sets defined by triangular membership functions. In the following discussion, note that the core of a fuzzy set is a singleton and is equal to its reference point when the fuzzy set is defined by a triangular membership function. Because a triangular membership function can be parameterized by its core and support, α-GEMII deduces only the core and support of the consequence by the effective use of its α-cut based scheme [5]:

Deduction of cores: The core $y_{\tilde{Q}^{\mathrm{c}}}$ of $\tilde{Q}$ is deduced by

$$y_{\tilde{Q}^{\mathrm{c}}}=M(\{y_{Q_j^{\mathrm{c}}},\tilde{p}_j\};1). \qquad (15)$$

Here, $y_{Q_j^{\mathrm{c}}}$ denotes the core of Q_j. Moreover, $\tilde{p}_j$ represents the compatibility degree between $\tilde{P}$ and P_j.

Deduction of supports: The least upper bound $y^{u}_{\underline{\tilde{Q}}}$ and greatest lower bound $y^{\ell}_{\underline{\tilde{Q}}}$ of the support $\underline{\tilde{Q}}$ of $\tilde{Q}$ are deduced by

$$y^{u}_{\underline{\tilde{Q}}}=M(\{y^{u}_{\underline{Q}_j}+(1-y_{\tilde{Q}^{\mathrm{c}}}),\tilde{p}_j\};\ \underline{\omega})-(1-y_{\tilde{Q}^{\mathrm{c}}}), \qquad (16)$$

$$y^{\ell}_{\underline{\tilde{Q}}}=\bar{M}(\{y^{\ell}_{\underline{Q}_j}-y_{\tilde{Q}^{\mathrm{c}}},\ \tilde{p}_j\};\ \underline{\omega})+y_{\tilde{Q}^{\mathrm{c}}}, \qquad (17)$$

where $y^{u}_{\underline{Q}_j}$ and $y^{\ell}_{\underline{Q}_j}$ represent the least upper and the greatest lower bounds of the support $\underline{Q}_j$ of Q_j, respectively. The function $\bar{M}(\{x_j,p_j\};\omega)$ is defined by

$$\overline{M}(\{x_j, p_j\}; \omega) = 1 - M(\{1-x_j,\ p_j\}; \omega). \quad (18)$$

The operation with $\overline{M}$ is called *the dual operation of* M in this paper. The value of $\underline{\omega}$ determines the support width of $\tilde{Q}$. In the general form of α-GEMII, the value of $\underline{\omega}$ is automatically controlled on the basis of the relation between $\tilde{P}$ and P_j $(j = 1, 2, \ldots, n)$ in fuzzy constraints [3]. In this paper, the value of $\underline{\omega}$ is fixed to 1 for applying α-GEMII to interval prediction. Thereby, $\underline{\tilde{Q}}$ is determined only by $\underline{Q}_j$ $(j = 1, 2, \ldots, n)$ and it can easily be used as a prediction interval after optimizing $\underline{Q}_j$ $(j = 1, 2, \ldots, n)$. When $\underline{\omega} = 1$, Eqs. (19) and (20) respectively turn to

$$y^u_{\underline{\tilde{Q}}} = M(\{y^u_{\underline{Q}_j}, \tilde{p}_j\}; 1), \quad (19)$$

$$y^\ell_{\underline{\tilde{Q}}} = M(\{y^\ell_{\underline{Q}_j}, \tilde{p}_j\}; 1). \quad (20)$$

The scheme of α-GEMII for triangular membership functions is detailed in [2–5].

4.2. Coverage-Width-Based Criterion

Prediction intervals are required to be estimated so that they include observed data $\hat{y}_k$ $(k = 1, 2, \ldots, n_d)$ while their widths are reduced to a great extent. In order to evaluate the prediction intervals, CWC (coverage-width-based criterion) was proposed in [6]. This study utilizes CWC for optimizing $\underline{Q}_j$ $(j = 1, 2, \ldots, n)$ for interval prediction with α-GEMII. The support $\underline{\tilde{Q}}$ of $\tilde{Q}$ is treated as a prediction interval.

The numerical index E_{CWC} of CWC is defined by the following equations:

$$E_{\text{CWC}} = \frac{w}{\sigma(c, \eta, a)}, \quad (21)$$

$$w = \frac{1}{n_d W} \sum_{k=1}^{n_d} (\tilde{y}^u_k - \tilde{y}^\ell_k), \quad (22)$$

$$c = \frac{1}{n_d} \sum_{k=1}^{n_d} c_k, \quad (23)$$

$$\sigma(c, \eta, a) = \frac{1}{1 + e^{-\eta(c-a)}}. \quad (24)$$

In Eq. (22), $\tilde{y}^u_k$ and $\tilde{y}^\ell_k$ respectively denote the least upper and the greatest lower bounds of the prediction interval. The symbol W represents the width of Y that is to be the space for $\hat{y}_k$. Eq. (22) gives the normalized mean prediction interval width (NMPIW). The symbol c_k in Eq. (23) is defined by

$$c_k = \begin{cases} 1, & \hat{y}_k \in [\tilde{y}^\ell_k, \tilde{y}^u_k], \\ 0, & \text{otherwise}. \end{cases} \quad (25)$$

The value of c in Eq. (23) presents the probability of the observed data included in the prediction intervals. It is called the prediction-interval coverage probability (PICP). Because Eq. (23) includes the operation for counting the number of the observed data in the prediction interval, it is a step function and therefore E_{CWC} is not differentiable.

4.3. An Artificial Immune System for Interval Prediction with α-FUZZI-EX learning

For constructing prediction intervals, fuzzy rules are optimized with an artificial immune system (AIS) [7, 8]. In the following, an AIS is specialized to the case where prediction intervals are constructed by α-GEMII on the basis of α-FUZZI-EX learning. For basic studies, Q_j $(j = 1, 2, \ldots, n)$ are defined by symmetric fuzzy sets. Suppose that the values of $y_{Q^c_j}$ $(j = 1, 2, \ldots, n)$ are precisely optimized in advance and only the support widths of Q_j $(j = 1, 2, \ldots, n)$ are to be tuned with the AIS for interval prediction. Thereby, it is easier to confirm the effectiveness of α-FUZZI-EX learning with CWC because the effect of the quality in optimizing the cores is eliminated.

In the AIS, antibodies Ab_{ji} $(j = 1, 2, \ldots, n;\ i = 1, 2, \ldots, n_{Ab})$ are the candidates of the support width of Q_j. As the fuzzy rules are optimized independently of each other in the same way, the process of the AIS for tuning the support width of Q_j is representatively presented in the following:

Step 1: Initialize Ab_{ji} $(i = 1, 2, \ldots, n_{Ab})$ with random numbers.

Step 2: Evaluate Ab_{ji} $(i = 1, 2, \ldots, n_{Ab})$ using E_{CWC}.

Step 3: Clone all antibodies Ab_{ji} $(i = 1, 2, \ldots, n_{Ab})$. The clones of Ab_{ji} are represented by C_{jik} $(k = 1, 2, \ldots, n_c)$.

Step 4: Mutate C_{jik} $(i = 1, 2, \ldots, n_{Ab};\ k = 1, 2, \ldots, n_c)$ by adding a random number whose magnitude is proportional to the value of E_{CWC} of Ab_{ji}. The mutation is performed so as to make the support of the consequent fuzzy set included in Y. The probabilities of making the support width of the consequent fuzzy set larger and smaller are equal.

Step 5: Evaluate C_{jik} $(i = 1, 2, \ldots, n_{Ab};\ k = 1, 2, \ldots, n_c)$ using E_{CWC}.

Step 6: Select n_{Ab} candidates with high performance in the evaluation by E_{CWC} from Ab_{jk} and C_{jik} $(i = 1, 2, \ldots, n_{Ab};\ k = 1, 2, \ldots, n_c)$.

Step 7: Replace the antibodies Ab_{ji} $(i = 1, 2, \ldots, n_{Ab})$ by the candidates selected in Step 6.

Step 8: Return to Step 3 if the predetermined termination condition is not satisfied; otherwise finish the process.

5. Simulations: Interval Prediction by Using α-FUZZI-EX Learning

Numerical results are presented to demonstrate the performance of α-FUZZI-EX learning via its application to interval prediction. In order to clarify the effectiveness in contrast to conventional methods, a non-differentiable function for evaluating fuzzy rules is adopted and facts are given by fuzzy sets (non-singletons).

5.1. Simulation Conditions

The simulations are performed under the following conditions:

(i) The number n of fuzzy rules is 21. The antecedent fuzzy set P_j and the consequent fuzzy set Q_j of R_j are symmetric. They are defined by triangular membership functions that satisfy Eq. (12). The cores x^c_j $(j = 1, 2, \ldots, n)$ of P_j $(j = 1, 2, \ldots, n)$ are respectively placed at $x = 0.05(j-1)$ $(j = 1, 2, \ldots, n)$. In the simulations, P_j $(j = 1, 2, \ldots, n)$ are fixed and are not adjusted for basic studies. Only the support widths of

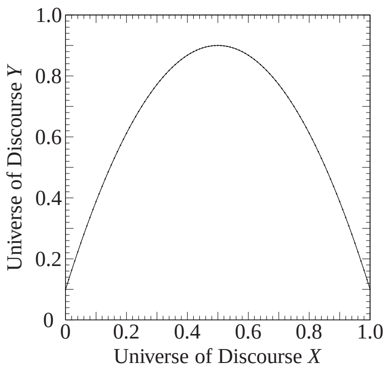


Fig. 1. Function $q(x)$

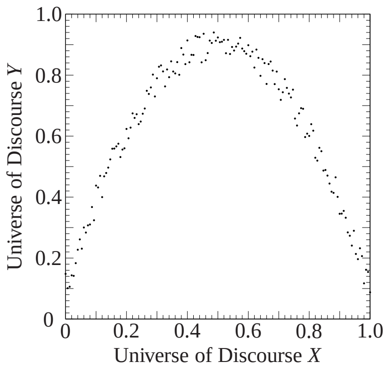


Fig. 2. Numerical learning data.

Q_j $(j = 1,2,\ldots,n)$ are optimized by using α-FUZZI-EX learning.

(ii) The input–output pairs $(\hat{x}_k, \hat{y}_k)$ $(k = 1,2,\ldots,n_d)$ of numerical learning data are generated by the following equations:

$$\hat{y}_k = q(\hat{x}_k) + r_k, \quad k = 1,2,\ldots,n_d, \tag{26}$$

$$q(x) = -3.2(x-0.5)^2 + 0.9, \tag{27}$$

where r_k denotes the additional noise. **Fig. 1** depicts the function $q(x)$. In order to confirm that α-FUZZI-EX learning leads to the proper optimization of fuzzy rules for interval prediction, $q(x)$ is defined so as to have some shapes. As can be found in **Fig. 1**, $q(x)$ has convex, increasing, and decreasing parts. The number n_d of the numerical learning data is 151. The values of $\hat{x}_k$ $(k = 1,2,\ldots,n_d)$ are placed at equal intervals in $[0,1]$. The value of r_k is given by a uniform random number for a feasibility study and is generated in $[-0.05, 0.05]$. **Fig. 2** shows the numerical learning data generated under the above-mentioned conditions. In order to confirm the proper convergence in fuzzy rules optimization by α-FUZZI-EX learning with facts given by fuzzy sets (non-singletons), the input $\hat{x}_k$ of the numerical learning data is transformed to the fuzzy set (non-singleton) $\hat{P}_k$. Here, $\hat{P}_k$ $(k = 1,2,\ldots,n_d)$ are symmetric and their membership functions are triangular. Their cores are set to $\hat{x}_k$ $(k = 1,2,\ldots,n_d)$, respectively. The support width of $\hat{P}_k$ is 0.05.

(iii) As described in Section 4.3, the values of $y_{Q_j^c}$ $(j = 1,2,\ldots,n)$ are set by $q(x_j^c)$ $(j = 1,2,\ldots,n)$, respectively. The widths of $\underline{Q}_j$ $(j = 1,2,\ldots,n)$ are adjusted independently of each other by using the AIS. The number n_{Ab} of antibodies for each consequent fuzzy set is 5. The antibodies are initialized by uniform random numbers in $[0, 0.2]$. The number n_c of clones of each antibody is 5. The antibodies and the clones are evaluated by using E_{CWC}. The values of W, η, and a are set to 1.0, 20.0, and 1.0, respectively. The mutation process for each clone is shown in the following: The values of E_{CWC} for evaluating Ab_{ji} and C_{jik} $(i = 1,2,\ldots,n_{Ab};\ k = 1,2,\ldots,n_c)$ are normalized with their maximum value in optimizing the width of $\underline{Q}_j$. The normalized value is denoted by $\acute{E}_{\mathrm{CWC}}$. Each clone is mutated in the way that the value of $\acute{E}_{\mathrm{CWC}} r_{\mathrm{mt}}$ is added to or subtracted from the least upper or greatest lower bounds of $\underline{Q}_j$, where r_{mt} is a uniform random number in $[0, r^*_{\mathrm{mt}}]$. The value of r^*_{mt} is adjusted so that the same process as the mutation with r^*_{mt} on behalf of $\acute{E}_{\mathrm{CWC}} r_{\mathrm{mt}}$ makes the least upper bound of $\underline{Q}_j$ equal to 1 and/or its greatest lower bound equal to 0 under the condition that Q_j is symmetric in the range $[0,1]$ of Y. Thereby, the mutation is performed so as to make $\underline{Q}_j$ included in the range $[0,1]$ of Y. The AIS is iteratively performed until 500 generations.

5.2. Numerical Results

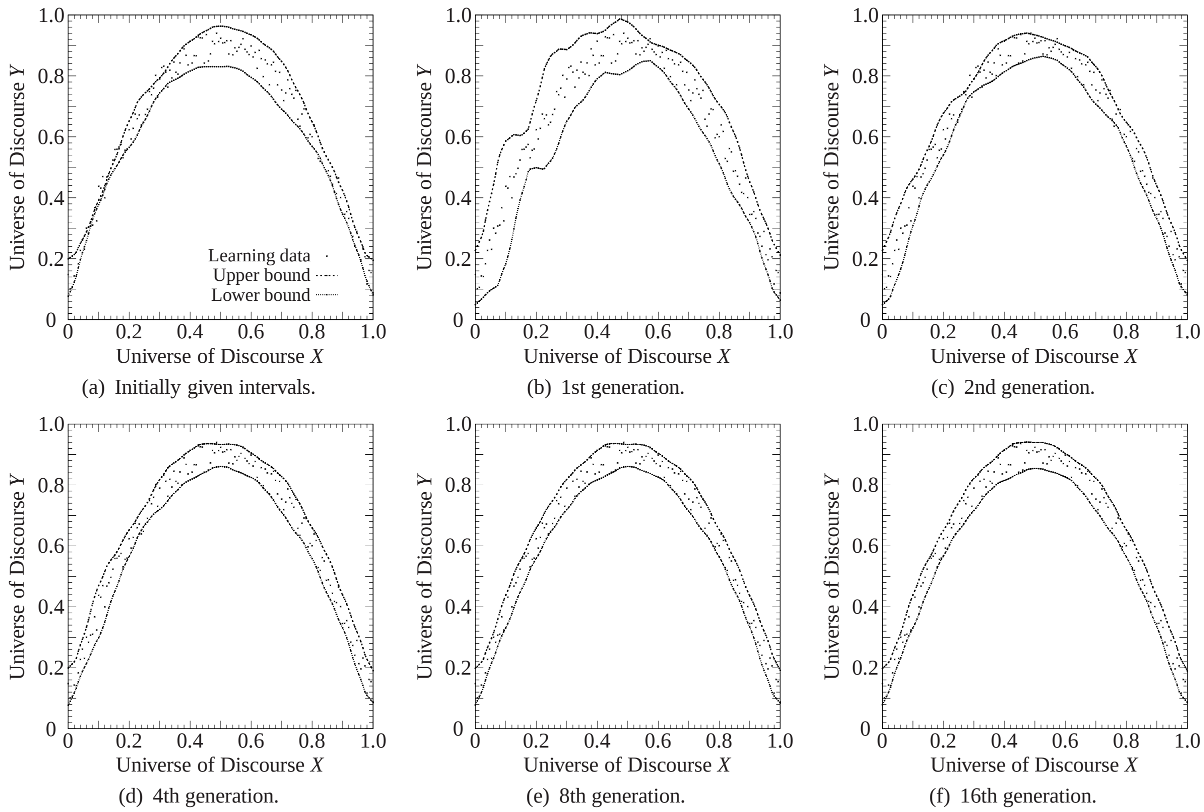
The prediction intervals in some generations are shown in **Fig. 3**. They are deduced by α-GEMII with $\hat{P}_k$ $(k = 1,2,\ldots,n_d)$ given as facts. **Fig. 3(a)** depicts the prediction intervals deduced with the initial fuzzy rules that are generated by using random numbers as described in Section 5.1. **Fig. 3(b)** shows the prediction intervals deduced in the first generation. Subsequently, **Figs. 3(c)-(f)** show the prediction intervals deduced in the other generations. The number of generations is indicated below each of the figures. As can be found in **Figs. 3(b)-(f)**, α-GEMII deduces the prediction intervals so that the numerical learning data are included in the intervals while the widths of the intervals are reduced to a great extent along with the generations.

The transitional changes of E_{CWC} in the learning are depicted in **Fig. 4**. As can be found from the figure, the values of E_{CWC} converge to small values along with the generations. Although **Fig. 4** presents the transitional changes of CWC until 30 generations, it is confirmed that the values of E_{CWC} further decrease in 500 generations.

From the above discussions, α-FUZZI-EX learning makes fuzzy rules properly converge. Therefore, it is found to be effective even when facts are given by fuzzy sets (non-singletons).

6. Conclusion

A method has been proposed for evaluating fuzzy rules independently of each other. It is derived by extending α-FUZZI-ES so as to cope with facts given by fuzzy sets (non-singletons). The proposed method has been named *α-FUZZI-EX* in this paper. This paper has further proposed a method for fuzzy rules learning on the basis of α-FUZZI-EX. The method provides a way to optimize fuzzy rules independently of each other in parallel. It has been



(a) Initially given intervals. (b) 1st generation. (c) 2nd generation. (d) 4th generation. (e) 8th generation. (f) 16th generation.

Fig. 3. Transitional changes in interval prediction by using α-FUZZI-EX learning and α-GEMII.

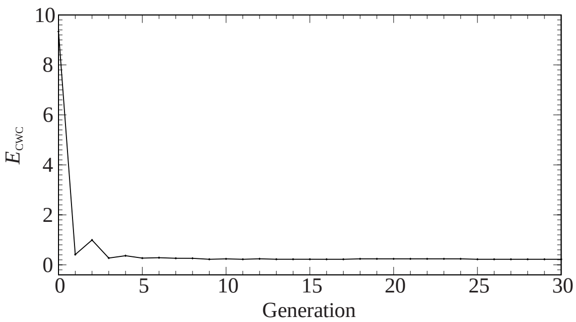


Fig. 4. Transitional changes of CWC.

named α-*FUZZI-EX learning* in this paper. α-FUZZI-EX learning attains fast fuzzy-rules learning by parallel processing hardware including GPGPUs and many-core CPUs. α-FUZZI-EX learning is effective especially when evaluation functions for fuzzy rules learning are not differentiable and then the derivative-based methods cannot be applied to the independent optimization of each fuzzy rule. As each fuzzy rule can evaluate itself and can self-govern in its optimization with α-FUZZI-EX learning, it is useful to apply evolutionary algorithms for optimizing fuzzy rules independently of each other. Such scheme of α-FUZZI-EX learning also contributes to reducing the dimensionality of the solution space for finding the optimal fuzzy rules. It makes the convergence speed fast in fuzzy rules learning.

In simulations, α-FUZZI-EX learning has been applied to interval prediction. In order to show its effectiveness, it optimizes each fuzzy rule with CWC that is represented by a non-differentiable function. Facts are given by fuzzy sets (non-singletons) to demonstrate the performance of α-FUZZI-EX learning in contrast to α-FUZZI-ES learning. Numerical results have shown that α-FUZZI-EX learning leads to proper convergence in constructing the prediction intervals.

References:

[1] K. Uehara and K. Hirota, "Independent evaluations of each fuzzy rule for derivative-free optimization of fuzzy systems: A feasibility study toward fast fuzzy-rule learning," Proc. of the 6th Int. Workshop on Advanced Computational Intelligence and Intelligent Informatics (IWACIII 2019), SUN4-C1, pp. 1-8, 2019.

[2] K. Uehara, T. Koyama, and K. Hirota, "Fuzzy inference with schemes for guaranteeing convexity and symmetricity in consequences based on α-cuts," J. of Adv. Comput. Intell. Intell. Inform. (JACIII), Vol. 13, No. 2, pp. 135-149, 2009.

[3] K. Uehara, T. Koyama, and K. Hirota, "Inference with governing schemes for propagation of fuzzy convex constraints based on α-cuts," J. of Adv. Comput. Intell. Intell. Inform. (JACIII), Vol. 13, No. 3, pp. 321-330, 2009.

[4] K. Uehara and K. Hirota, "Noise reduction with inference based on fuzzy rule interpolation at an infinite number of activating points: Toward fuzzy rule learning in a unified inference platform," J. of Adv. Comput. Intell. Intell. Inform., Vol. 22, No. 6, pp. 883-899, 2018.

[5] K. Uehara and K. Hirota, "Fuzzy inference based on α-cuts and generalized mean: Relations between the methods in its family and their unified platform," J. of Adv. Comput. Intell. Intell. Inform. (JACIII), Vol. 21, No. 4, pp. 597-615, 2017.

[6] A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction interval construction and optimization for adaptive neurofuzzy inference systems," IEEE Trans. Fuzzy Syst., vol. 19, no. 5, pp. 983-988, October 2011.

[7] L. N. de Castro and F. J. von Zuben, "Learning and optimization using the clonal selection principle," IEEE Trans. Evol. Comput., Vol. 6, No. 3, pp. 239-251, 2002.

[8] K. M. Woldemariam and G. G. Yen, "Vaccine-enhanced artificial immune system for multimodal function optimization," IEEE Trans. Syst. Man Cybern.—Part B: Cybernetics, Vol. 40, No. 1, pp. 218-228, 2010.