Marian B. GORZAŁCZANY, Filip RUDZIŃSKI
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING, KIELCE UNIVERSITY OF TECHNOLOGY
Al. 1000-lecia Państwa Polskiego 7, 25-314 Kielce

Measurement data in genetic fuzzy modelling of dynamic systems

Prof. Marian B. GORZAŁCZANY

In 1978 he received the M.Sc. degree in electronics from Warsaw Univ. of Technology, in 1983 and 1994 – the Ph.D. and D.Sc. degrees in computer science from Poznań Univ. of Technology and in 2003 – the professor title. In 1990-1992 he was working at Univ. of Saskatchewan and Univ. of Guelph, Canada. He is the author/co-author of more than 80 refereed scientific works including the monograph published by Springer-Verlag. The main area of scientific interests is designing and application of computational intelligence systems.

e-mail: m.b.gorzalczany@tu.kielce.pl

Ph.D. eng. Filip RUDZIŃSKI

He graduated from Department of Electrical and Computer Engineering, Kielce University of Technology (specialization: electrical engineering) in 2000. He received his Ph.D. degree in computer science from Systems Research Institute of Polish Academy of Sciences in 2006. The area of his scientific interests covers artificial intelligence, including neural networks, fuzzy systems and genetic algorithms.

e-mail: f.rudzinski@tu.kielce.pl

Abstract

The paper presents a genetic fuzzy rule-based approach to the modelling of complex dynamic systems and processes using measurement data that describe their behaviour. The application of the proposed technique to modelling an industrial gas furnace system (the so-called Box-Jenkins benchmark) using measurement data available from the repository at the University of Wisconsin at Madison (http://www.stat.wisc.edu/~reinsel/bjr-data) is also presented in the paper.

Keywords: computational intelligence, fuzzy systems, genetic algorithms, genetic fuzzy systems, measurement data.

Dane pomiarowe w genetyczno-rozmytym modelowaniu systemów dynamicznych

Streszczenie

Artykuł prezentuje podejście genetyczno-rozmyte do modelowania (z wykorzystaniem zestawów reguł rozmytych) złożonych, dynamicznych systemów i procesów na bazie danych pomiarowych opisujących ich zachowanie. Najpierw sformułowany został problem budowy modeli (w formie zestawów reguł rozmytych) systemów dynamicznych z wykorzystaniem danych opisujących ich zachowanie. Następnie przedstawiono proces syntezy reguł rozmytych z danych z wykorzystaniem zaproponowanego przez autorów zmodyfikowanego podejścia typu Pittsburgh z obszaru algorytmów genetycznych. Z kolei, przedstawiono zastosowanie proponowanej techniki do modelowania systemu przemysłowego pieca gazowego (tzw. benchmark Box'a-Jenkins'a) z wykorzystaniem danych pomiarowych dostępnych w repozytorium Uniwersytetu Wisconsin w Madison, USA (http://www.stat.wisc.edu/~reinsel/bjr-data). Uzyskany model, w formie zestawu reguł rozmytych, przetestowano w trybie pracy predyktora jednokrokowego oraz wielokrokowego (na pełnym horyzoncie symulacji). Dokonano również analizy zależności pomiędzy dokładnością a przejrzystością (mierzoną liczbą reguł) modelu oraz przetestowano model ze zredukowaną bazą reguł.

Słowa kluczowe: inteligencja obliczeniowa, systemy rozmyte, algorytmy genetyczne, systemy genetyczno-rozmyte, dane pomiarowe.

1. Introduction

Measurement data can be a very useful and comprehensive source of information regarding the behaviour of many complex systems and processes. On the other hand, the so-called computational intelligence (henceforward: CI) systems, cf. [1], [4], based on synergistic combinations of such techniques as artificial neural networks, fuzzy logic, evolutionary computations, rough sets, etc., provide powerful tools for intelligent data analysis and discovery of the knowledge (in the form of major trends, patterns, rule-based decision mechanisms, etc.) in the data. Therefore, the CI techniques can be effectively used in measurement-data-based modelling of dynamic systems. In particular, fuzzy rule-based models designed with the use of measurement data are of special interest here. Hybrid CI approaches that combine genetic algorithms and fuzzy logic – referred to as genetic fuzzy systems, cf. [3] – are particularly effective in fuzzy rule-based modelling since they are the tools for the optimization of both the structure and parameters of the fuzzy models.

This paper briefly presents a genetic fuzzy rule-based modelling technique that is a generalization of the genetic fuzzy classifier introduced by the same authors in [5]. First, the problem of the fuzzy rule-based model design of dynamic systems using data is formulated. Then, a genetic learning of fuzzy rules from data based on a modification of the so-called Pittsburgh approach (cf. [5]) is outlined. Finally, the application of the proposed technique to modelling an industrial gas furnace system using measurement data (the so-called Box-Jenkins benchmark [2]) is presented. The data are available from the repository at the University of Wisconsin at Madison (http://www.stat.wisc.edu/~reinsel/ bjr-data).

2. Designing a fuzzy rule-based model of a dynamic system from data

A dynamic system with r inputs $u_1, u_2, ..., u_r$ ($u_c \in U_c$, $c = 1,2,...,r$) and s outputs $z_1, z_2, ..., z_s$ ($z_d \in Z_d$, $d = 1,2,...,s$) is considered. Assume that the behaviour of the system is described by T input-output data samples:

$$D = \{\boldsymbol{u}_t', \boldsymbol{z}_t'\}_{t=1}^{T} \ , \qquad (1)$$

where $\boldsymbol{u}_t' = (u_{1t}', u_{2t}', ..., u_{rt}') \in \boldsymbol{U} = U_1 \times U_2 \times ... \times U_r$ and $\boldsymbol{z}_t' = (z_{1t}', z_{2t}', ..., z_{st}') \in \boldsymbol{Z} = Z_1 \times Z_2 \times ... \times Z_s$ ($\times$ stands for Cartesian product of ordinary sets). u_{ct}' and z_{dt}' are the data describing the c-th input ($c = 1,2,...,r$) and d-th output ($d = 1,2,...,s$) of the system in discrete time instant t. Due to system dynamics, the particular data records in (1) are interrelated. For instance, for a dynamic system with one input u and one output z its dynamics can be described, in general, by the following formula:

$$z_t = f(u_t, u_{t-1}, ..., u_{t-M}, z_{t-1}, z_{t-2}, ..., z_{t-N}) \ , \ M \geq 0 \ , \ N \geq 1 \ . \qquad (2)$$

Only in the case of a static system, formula (2) reduces to $z_t = f(u_t)$. Expression (2) can be easily generalized for the case of the system with r inputs and s outputs.

The essential stage of fuzzy rule-based model design consists in determination of the model structure in terms of its inputs and outputs similarly as in (2). It is an approximation of the system dynamics. Let us assume now that such a structure of the model has been determined and the model has n inputs $x_1, x_2, ..., x_n$, $x_i \in X_i$ ($n \geq r$), and m outputs $y_1, y_2, ..., y_m$, $y_j \in Y_j$ (usually $m = s$ and $y_j = z_j$). For instance, for a single input and single output system with the dynamics (2), $x_1 = u_t$, $x_2 = u_{t-1}, \ldots$, $x_{M+1} = u_{t-M}$, $x_{M+2} = z_{t-1}$, $x_{M+3} = z_{t-2}, \ldots$, $x_n = x_{M+N+1} = z_{t-N}$, and $y_1 = y = z_t$. The initial description (1) of the dynamic system has to be now reedited to the "static" form (according to the model structure):

$$L = \{\boldsymbol{x}'_k, \boldsymbol{y}'_k\}_{k=1}^{K}, \quad (3)$$

where $\boldsymbol{x}'_k = (x'_{1k}, x'_{2k}, \ldots, x'_{nk}) \in \boldsymbol{X} = X_1 \times X_2 \times \ldots \times X_n$, $\boldsymbol{y}'_k = (y'_{1k}, y'_{2k}, \ldots, y'_{mk}) \in \boldsymbol{Y} = Y_1 \times Y_2 \times \ldots \times Y_m$, k is the number of model input-output static data pattern and K is the overall number of such data patterns. Data (3) are referred to as the learning data.

Fuzzy rules that will be synthesized from data (3) (for the case of single output system, that is, $m = 1$ and $y_1 = y$) by the proposed later in the paper genetic technique have the following form (when all n inputs are involved):

IF (x_1 is A_{1r}) AND … AND (x_n is A_{nr}) THEN (y is B_r), (4)

where $A_{ir} \in F(X_i)$, $i = 1,2,\ldots,n$, and $B_r \in F(Y)$ are the S-, M-, or L-type fuzzy sets (see [5] for details) representing verbal terms Small, Medium and Large, respectively, in the *r*-th fuzzy rule, $r = 1,2,\ldots,R$. $F(X_i)$ and $F(Y)$ denote families of all fuzzy sets defined in the universes X_i and Y, respectively.

3. Genetic learning of fuzzy rules from data

In the course of learning fuzzy rules (4) from data by means of genetic algorithm, a rule base and a data base are being processed. No special coding of fuzzy rules has been used. The rule base simply contains the information about the present structure of particular rules indicating which antecedents are presently in use (they can be switched on and off). The data base contains decimal representations of the parameters of membership functions of particular fuzzy antecedents and consequences that occur in rules (4). An essential role is played by the proposed non-binary crossover and mutation operators. They have been defined separately for the rule base and data base processing.

The crossover operator defined for the rule base transformation processes two individuals (two rule bases) by executing one of five randomly selected sub-operators:

C1 (labelled as *MultiFuzzyRuleExchange*). Let R_1 denote the number of fuzzy rules in the first rule base and R_2 – in the second one. C1 operates in two stages. In the first stage, for the r-th fuzzy rule in both rule bases, $r = 1,2,\ldots,\min(R_1, R_2)$, the so-called *random-switch* condition (equivalent to the random selection of 1 from the set {0, 1}) is checked. If this condition is fulfilled, the r-th fuzzy rules from both rule bases are exchanged. The second stage deals with the remaining fuzzy rules of the larger rule base. Each of these rules, assuming that the *random-switch* condition is fulfilled, is moved to the smaller rule base.

C2 (labelled as *SingleFuzzyRuleExchange*). It operates in an analogous way as C1 but the activities of C1 are performed unconditionally only once for randomly selected r-th location in the larger rule base.

C3 (labelled as *EvenFuzzySetExchange*). For the r-th fuzzy rule in both rule bases, $r = 1,2,\ldots,\min(R_1, R_2)$, the *random-switch* condition is checked. If it is fulfilled then – for the i-th input attribute, $i = 1,2,\ldots,n$ and for the output attribute – the *random-switch* condition is run independently again. If this time the condition is fulfilled the fuzzy sets describing a given input/output attribute in both rule bases are exchanged.

C4 (labelled as *MultiFuzzySetExchange*). It operates analogously as C3 but the activities of C3 are performed unconditionally only once for randomly selected r-th location in both rule bases ($r \in \{1,2,\ldots,\min(R_1, R_2)\}$).

C5 (labelled as *SingleFuzzySetExchange*). It operates analogously as C4 but the activities of C4 are performed unconditionally only once for randomly selected i-th input attribute ($i \in \{1,2,\ldots,n\}$) or output attribute.

Due to a limited space of this publication, the crossover operator for the data base processing as well as the mutation operators for both the rule base and data base processing cannot be presented here; see [5] for details.

The fitness function ff for a multi input and single output system has been defined as follows: $ff = const. - Q$, where *const.* is a constant value selected in such a way that $ff > 0$ and Q is the cost function (a normalized root mean squared error Q_{rmse}):

$$Q = \frac{1}{y_{\max} - y_{\min}} Q_{rmse} = \frac{1}{y_{\max} - y_{\min}} \sqrt{\frac{1}{K} \sum_{k=1}^{K} (y'_k - y_k^0)^2}, \quad (5)$$

where K is the number of the learning samples in (3), $y_{\min}$ and $y_{\max}$ determine the range of changes of the system output $y = z_t$, y'_k is the desired model response for the *k*-th learning data sample of (3), and y_k^0 is the actual response of the model for that data sample.

4. Fuzzy rule-based modelling of an industrial gas furnace system

As an example of the application of the proposed genetic fuzzy technique, the fuzzy rule-based modelling of an industrial gas furnace system using the so-called Box-Jenkins' measurement data [2] will be briefly discussed. These data – available from the repository at the University of Wisconsin at Madison (http://www.stat.wisc.edu/~reinsel/bjr-data) – have become a benchmark in the assessment of new identification and modelling techniques. They consist of 296 pairs of the gas flow rate (input u_t) measured in ft^3/min and the concentration of CO_2 in the exhaust gas (output z_t) expressed in %. The sampling period is equal to 9 sec. Therefore, it is a single input and single output dynamic process described by data (1) with $r = 1$, $s = 1$ and $T = 296$. Applying two-input one-output structure of the model, the best approximation of its dynamics is described by the formula (see e.g. [4]):

$$z_t = f(u_{t-4},\ z_{t-1}). \quad (6)$$

In such a case, the original Box-Jenkins' data must be reedited from the collection of input-output pairs (1) $(u'_t,\ z'_t)$ to the collection of input-output triplets (3) $((u'_{t-4},\ z'_{t-1}),\ z'_t) = ((x'_1,\ x'_2),\ y')$; there are $K = 294$ such triplets (they will be used as the learning data). Fuzzy rules that will be discovered in the learning data have the general form of (4) with $n = 2$. However, the rules with single antecedent (either $x_1 = u_{t-4}$ or $x_2 = z_{t-1}$) can also be discovered.

In the reported experiment, the genetic algorithm with the population of 100 individuals and the tournament selection method (with the number of individuals participating in the competition [8] equal to 2) supported by the elitist strategy as well as with crossover and mutation probabilities equal to 0.8 and 0.7, respectively, has been used. After the learning completion, a rule pruning technique has been applied, in order to obtain smaller and more transparent fuzzy rule base.

Fig. 1 presents the plot of the cost function Q (5) for the best, worst and average individuals versus the number of generations.

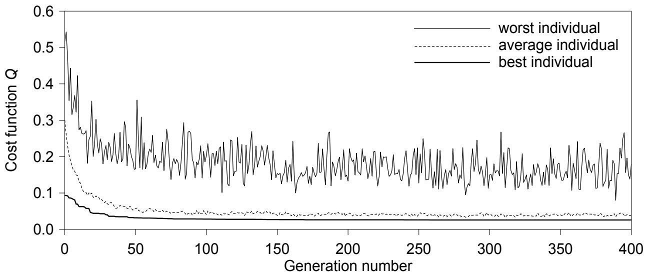

Fig. 1. Cost function Q (5) versus generation number plot
Rys. 1. Przebieg wartości funkcji kosztu Q (5) w trakcie procesu uczenia

Fig. 2 presents final shapes of the membership functions of fuzzy sets describing system's input u_t (*gas flow rate*) – Fig. 2a, and system's output z_t (*CO_2 concentration*) – Fig. 2b.

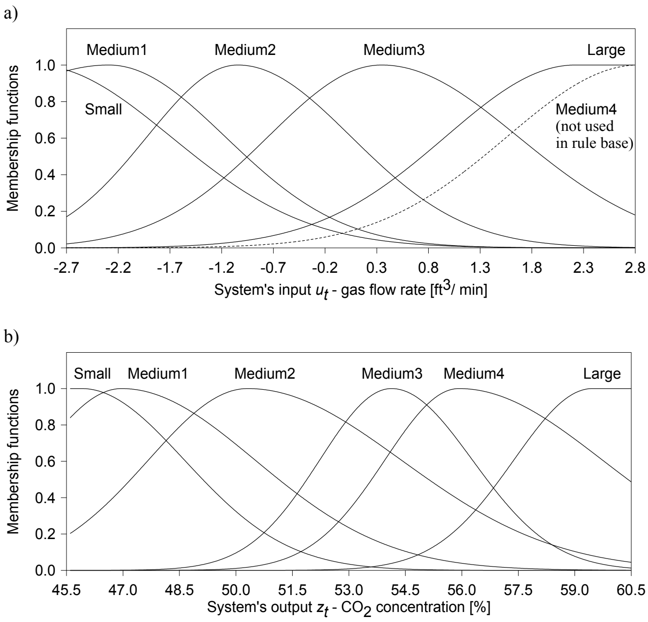

Fig. 2. Final shapes of membership functions of fuzzy sets describing system's input u_t (a) and system's output z_t (b)
Rys. 2. Końcowe kształty funkcji przynależności zbiorów rozmytych opisujących wejście systemu u_t (a) oraz jego wyjście z_t (b)

Table 1 presents the full rule base generated by the proposed approach. The last column and the last row in Table 1 represent fuzzy rules with single antecedent, $x_1 = u_{t-4}$ and $x_2 = z_{t-1}$, respectively.

Tab. 1. Final fuzzy rule base of the genetic fuzzy model – dark cells represent fuzzy rules removed from the rule base as a result of a pruning (see below)
Tab. 1. Końcowa baza reguł modelu genetyczno-rozmytego – zaciemnione komórki reprezentują reguły usunięte z bazy reguł w procesie tzw. pruningu (patrz poniżej)

$x_2=z_{t-1}$ / $x_1=u_{t-4}$	S	M1	M2	M3	M4	L	–
S	M2			M3			
M1					L		
M2			M2				M4
M3		M1	M2	M3		M4	
M4							
L	S						
–	S	S				L	

S=Small, M1=Medium1, M2=Medium2, M3=Medium3, M4=Medium4, L=Large $y=z_t$

The obtained fuzzy rule-based model will be tested in "OSA predictions" mode and in "AFT predictions" mode. OSA predictions stand for *O*ne-*S*tep-*A*head predictions. This means that the model – using $x'_1 = u'_{t-4}$ and $x'_2 = z'_{t-1}$ from the learning data set – produces a response $y^0 = z^0_t$ (one-step-ahead prediction) which, in turn, can be compared with the desired response $y' = z'_t$. A much tougher test of model's accuracy is its operation as the AFT ("*A*ll-*F*uture-*T*imes") predictor. In such a case, the model using $x'_1 = u'_{t-4}$ from the learning data set and $x'_2 = z^0_{t-1}$ generated by the model itself in the previous iteration, produces a response $y^0 = z^0_t$. The cumulation of errors associated with generation of $y^0 = z^0_t$ by the model in the consecutive iterations can cause – if the model's accuracy is not sufficiently high – that the model becomes more and more divergent with regard to the data. Fig. 3 presents the operation of the genetic fuzzy model with full rule base working as OSA predictor (a) and AFT predictor (b). The accuracy analysis of this model is summarized in Table 2.

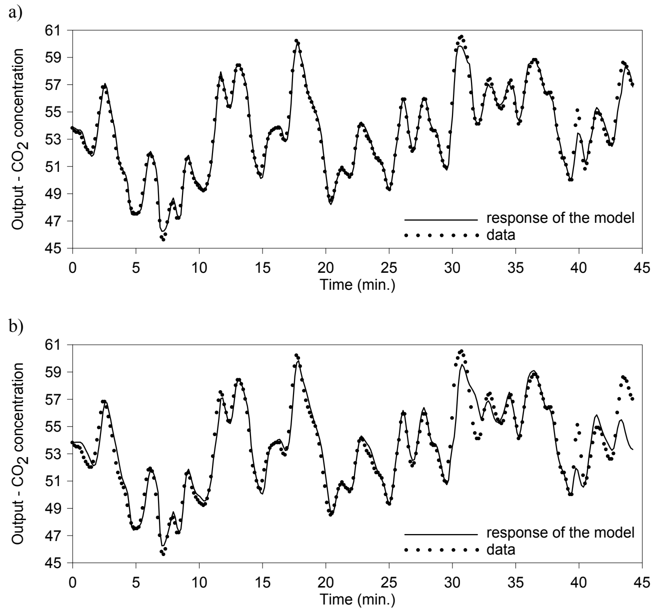

Fig. 3. Genetic fuzzy model with full rule base working as OSA predictor (a) and AFT predictor (b)
Rys. 3. Model genetyczno-rozmyty z pełną bazą reguł pracujący w trybie predykcji jednokrokowej (a) oraz wielokrokowej – na całym horyzoncie symulacji (b)

The last stage of the genetic fuzzy model design – the pruning of its structure – consists in gradual removing of "weaker", superfluous rules from the model's rule base (in order to increase its transparency), and analyzing how it affects the model's accuracy. This process is illustrated in Fig. 4. Q_{rmse} is the root mean squared error as in (5) calculated for the model (after the completion of the learning process) for different numbers of fuzzy rules in model's rule base.

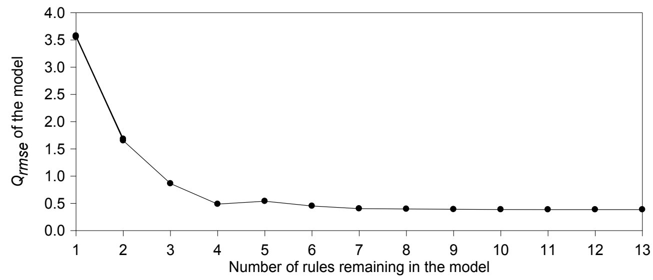

Fig. 4. Q_{rmse} accuracy criterion versus transparency criterion (number of rules in the model's rule base) for genetic fuzzy model
Rys. 4. Kryterium Q_{rmse} dokładności a kryterium przejrzystości (mierzonej liczbą reguł w bazie reguł) modelu genetyczno-rozmytego

After removing nine "weakest" rules (represented by dark cells in Table 1) from the full rule base, the model with reduced rule base – containing four "strongest" fuzzy rules – has been obtained. These rules can be expressed, according to formula (4), as follows:

$$\begin{aligned} &\text{IF } (u_{t-4} \text{ is } \mathsf{Medium2}) \text{ THEN } (z_t \text{ is } \mathsf{Medium4}) \\ &\text{IF } (u_{t-4} \text{ is } \mathsf{Medium3}) \text{ AND } (z_{t-1} \text{ is } \mathsf{Medium2}) \\ &\qquad\qquad\qquad\qquad\quad \text{THEN } (z_t \text{ is } \mathsf{Medium2}) \\ &\text{IF } (z_{t-1} \text{ is } \mathsf{Medium1}) \text{ THEN } (z_t \text{ is } \mathsf{Small}) \\ &\text{IF } (z_{t-1} \text{ is } \mathsf{Large}) \quad\; \text{THEN } (z_t \text{ is } \mathsf{Large}), \end{aligned} \qquad (7)$$

where the membership functions of fuzzy sets representing verbal terms Medium2 and Medium3 for u_t as well as Small, Medium1, Medium2, Medium4 and Large for z_t are presented in Figs. 2a and 2b, respectively.

The operation of the genetic fuzzy model with reduced rule base working as OSA predictor and AFT predictor is illustrated in Figs. 5a and 5b, respectively. Moreover, its accuracy analysis is presented in Table 2.

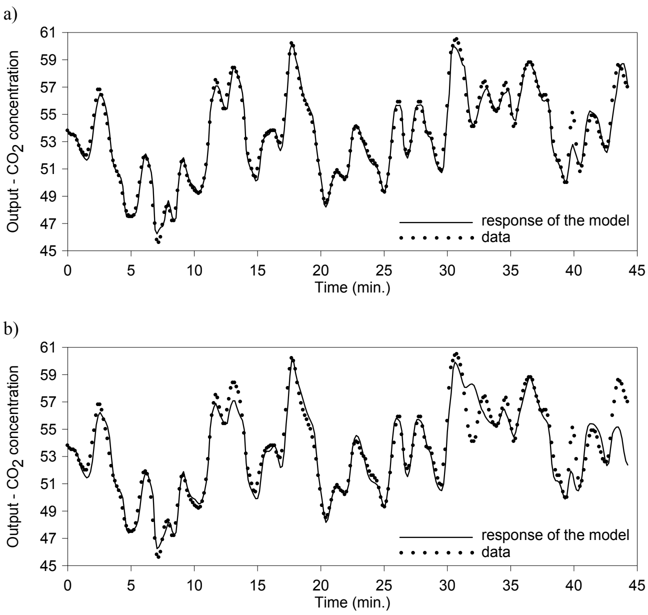

Fig. 5. Genetic fuzzy model with reduced rule base working as OSA predictor (a) and AFT predictor (b)

Rys. 5. Model genetyczno-rozmyty ze zredukowaną bazą reguł pracujący w trybie predykcji jednokrokowej (a) oraz wielokrokowej – na całym horyzoncie symulacji (b)

Tab. 2. Q_{rmse} accuracy of genetic fuzzy model with full and reduced rule bases working in OSA and AFT prediction modes

Tab. 2. Miara Q_{rmse} dokładności modelu genetyczno-rozmytego z pełną i zredukowaną bazą reguł pracującego w trybie predykcji jednokrokowej oraz wielokrokowej

	Full rule base	Reduced rule base
Q_{rmse} (OSA predictions)	0.385	0.487
Q_{rmse} (AFT predictions)	0.963	1.064
Number of rules in rule base	13	4

It seems that the genetic fuzzy model with the reduced rule base containing only four fuzzy rules (7) provides an excellent compromise between high transparency and interpretability of the model and, on the other hand, its high accuracy. The fuzzy rules describe, in a clear and easy-to-comprehend way, the dynamics that governs the functioning of the industrial gas furnace system. These rules also explain how that system works. On the other hand, as shown in Table 2, they also guarantee a very high accuracy (in terms of root mean squared error Q_{rmse}) of the fuzzy model. Especially, its operation in a very demanding AFT prediction mode (multi-step-ahead prediction on the whole simulation horizon) confirms its high accuracy.

5. Conclusions

The genetic fuzzy rule-based technique for modelling complex dynamic systems and processes using measurement data that describe their behaviour has been briefly presented in this paper. The proposed technique is a generalization of the genetic fuzzy classifier introduced by the same authors in [5].

Designing fuzzy rule-based models from input-output data sets can be treated as structure- and parameter-optimization problems. Therefore, only the global search methods in complex spaces can lead to optimal or sub-optimal solutions to the problems at hand. It seems that ideal methods that can provide such solutions are offered by evolutionary computations and, in particular, by genetic algorithms. That is the motivation and rationale behind the research that has been briefly reported here. The obtained results are very promising. As it has been shown, the proposed genetic fuzzy rule-based modelling technique is the solution characterized by both, high accuracy of the model and its high transparency, interpretability and ability to explain the generated actions. An alternative evolutionary approach (based on evolution strategies) to fuzzy rule based modelling of complex systems using input-output data has been briefly presented in [6] (in this volume). Moreover, a hybrid combination of fuzzy-rule systems and the so-called particle swarm optimization approach for optimizing sets of fuzzy control rules synthesized from control data has also been presented in [7] (in this volume).

6. References

[1] Bezdek J.C.: What is computational intelligence?, in J.M. Zurada, R.J. Marks II, C.J. Robinson (Eds.) Computation Intelligence: Imitating Life, IEEE Press, New York, 1994, pp. 1-12.

[2] Box G.E., Jenkins G.M.: Time Series Analysis: Forecasting and Control, Holden Day, San Francisco, 1970.

[3] Cordon O., Herrera F., Hoffmann F., Magdalena L.: Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases. World Scientific, Singapore, New Jersey, London, Hong Kong, 2001.

[4] Gorzałczany M.B.: Computational Intelligence Systems and Applications, Neuro-Fuzzy and Fuzzy Neural Synergisms. Physica-Verlag, Springer-Verlag Co., Heidelberg, New York, 2002.

[5] Gorzałczany M.B., Rudziński F.: A modified Pittsburgh approach to design a genetic fuzzy rule-based classifier from data, in L. Rutkowski, R. Scherer, R. Tadeusiewicz, L.A. Zadeh, J. Żurada (Eds.), Artificial Intelligence and Soft Computing – ICAISC 2010, Part I, Lecture Notes in Artificial Intelligence 6113, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 88-96.

[6] Gorzałczany M.B., Głuszek A.: Data-based fuzzy modelling of dynamic systems by means of evolution strategies. Pomiary Automatyka Kontrola, vol. 56, no. 12, 2010, pp. 1427 - 1429.

[7] Gorzałczany M.B., Głuszek A.: Designing fuzzy rule-based controllers from data using particle swarm optimization. Pomiary Automatyka Kontrola, vol. 56, no. 12, 2010, pp. 1424 - 1426.

[8] Michalewicz Z.: Genetic Algorithms + Data Structures = Evolution Programs. Springer-Verlag, Berlin, Heidelberg, New York, 1996.

otrzymano / received: 25.09.2010
przyjęto do druku / accepted: 01.11.2010 *artykuł recenzowany*