Complexity reduction of rule based models: a survey

Okyay Kaynak * Karel Jezernik ** and Ágnes Szeghegyi ***
* Dept. Electrical Eng. Bogazici Univ. Debek 80815 Istambul. Turkey, Kaynak@boun.edu.tr
** Karel Jezernik University of Maribor, Slovenia karel.jezernik@uni-mb.si
*** Budapest PolytechnicH-1081 Budapest Nepszinhaz u.8. Hungary

Abstract - This paper gives a survey of fuzzy rule base reduction methods. The research of complexity reduction methods are originated form two aspects depending of different design methodologies of the model. The first model design type comes from the original idea of Zadeh, it proposes models which are built based on expert knowledge, hence the rule base is set up manually. These models feature linguistic, and hence semantically interpretable fuzzy terms, and rules with fuzzy sets as consequents. However, it is often the case that the model contains redundant rules and/or variables, so there is a need to omit the redundancy of the model. Secondly, in the last decade data-driven fuzzy model design became more popular. This is partly due to fact that fuzzy models were found to be universal approximators, i.e. they are capable to approximate with arbitrary accuracy any continuous control function. For fitting the model the possible best to the approximated function, these models, usually having rules with consequents which are linear function of the inputs, use tremendously large number of rules, and do not take into account the complexity and interpretability of the model. This feature also emerged the issue of rule base reduction for such systems. The present paper aims at summarizing the efforts done on the complexity reduction field briefly.

I. Introduction

FUZZY MODELS describes systems by determining the relation between the prospective inputs and the output of the its model in the form of IF-THEN rules. The are different methodologies of model design, which can mainly divided into two groups: the semantic-driven modelling, and data-driven modelling.

The first model design type comes from the original idea of Zadeh [1], it proposes models which are built based on expert knowledge, hence the rule base is set up manually. In practice the Mamdani version [2] of fuzzy model became very popular, which worked in the projection spaces instead of the multidimensional product space, and hence required much less computational effort. Semantic-based models feature linguistic, and hence semantically interpretable fuzzy terms. The rules are usually of Mamdani type, i.e. with fuzzy sets in the consequents:

$$R_i: \text{ If } x_1 \text{ is } A_{i1} \text{ and} \ldots \text{and } x_n \text{ is } A_{in} \text{ then } y \text{ is } B_i.$$

However, it is often the case that the model contains redundant rules and/or variables, so there is a need to omit the redundancy of the model. In Section II we survey the methods proposed for complexity reduction of rule base designed by semantic-driven model.

In the last decade data-driven fuzzy model design became more popular. This is partly due to fact that fuzzy models were found to be universal approximators (see e.g. [3], [4], [5]), i.e. they are capable to approximate with arbitrary accuracy any continuous control function. To achieve the arbitrary accuracy in approximation requires usually unbounded number of rules, as it was pointed out in [6]. Moreover, if the number of rules are bounded the set of resulted models is almost discrete in the set of continuous functions [7], [8].

This feature urged researchers to search for practically feasible, but still sufficiently good approximative models for (control) function approximation. The models are mostly of Takagi–Sugeno–Kang (TSK) type [9], [10], where the rule consequents are (usually) linear function of the inputs:

$$R_i: \text{ If } x_1 \text{ is } A_{i1} \text{ and} \ldots \text{and } x_n \text{ is } A_{in} \text{ then } y_i = \mathbf{a_i x} + b_i$$

where $\mathbf{a}_i = [a_{i1}, \ldots, a_{in}]$ is the parameter vector, and $\mathbf{x} = [x_1, \ldots, x_n]^T$ is the input vector. Due to the non-fuzzy consequent this model is easier to handle by classical numerical in the following manner. The overall output is written as the weighted rule output of local models described by individual rules:

$$y = \sum_{i=1}^{M} w_i(\mathbf{x}) y_i \qquad (1)$$

0-7803-7280-8/02/$10.00 ©2002 IEEE

where $w_i(\mathbf{x})$ is the normalized firing strength of the ith rule. The equation (1) offers a way to handle to the problem as linear regression by classical numerical methods. The Section III surveys the appropriate numerical methods.

Before going into the details of complexity reduction we should remark here, that fuzzy models have originally exponential complexity [11]. The number of the rules is of order $O(T^k)$, where T is the maximum (or average) number of fuzzy terms in each dimension, and k is the number of dimensions. When searching for the relevant rules for a given input (observation), an exhaustive search among all the rules should be performed, therefore the exponential complexity is inevitable. In fact, there is two different ways of reducing the computational need of the model: the reduction of the number of terms, T, or the number of variables, k. As a third possibility, these techniques might be combined, resulting in methods decreasing both of the above quantities.

II. Reduction of semantic-based models

A. Fuzzy rule interpolation

Fuzzy rule interpolation was one of first approaches to reduce the complexity of fuzzy models. The first such method , proposed by Kóczy and Hirota [12], [13], is linear fuzzy rule interpolation. Its main idea is that particular fuzzy rules, which can be (approximately) substituted by linear interpolation of the neighbouring rules, can be omitted from the rule base. Hence, the linear section of the input-output function of the model needs only two rules two describe, the rules in-between them are not necessary. Therefore this method aims at reducing T in the equation of complexity.

However, the thinned out rule base usually does not satisfy the applicability criterion of rule matching based inference techniques (including Zadeh, Mamdani, and TSK type models), because the rule antecedents often do not give a full α-coverage of the input dimensions ($\alpha > 0$), i.e. the rule base is (α-) sparse. This means that these inference techniques should be replaced by a new approach. The KH linear rule interpolation is such an approach which is able to work on sparse rule bases. It determines the conclusion by its α-cuts in such a way that the ratio of distances among the conclusion and the consequents should be identical with the ones among observation and the antecedents for all important α-cuts. This fundamental equation of KH linear rule interpolation is an extension of the linear interpolation for rules based convex and normal fuzzy sets in accordance with the gradual semantic interpretation, proposed first by Dubois and Prade in 1992 [14], "the more similar is the observation to an antecedent the more similar the conclusion should be to the corresponding consequent of the given antecedent".

The linear interpolation can be applied if: 1. there exists an ordering among the fuzzy sets in each dimension, 2. the observation is located between two antecedents, and 3. all involved fuzzy sets are convex and normal (CNF).

The KH method has a significant drawback: it may produce conclusions which are not interpretable (abnormal) as fuzzy set. This problem urged several researchers to solve the problem by characterizing the abnormal situation, or by proposing conceptually new techniques. Among those here we mention the solid cutting based method [15], and the modified α-cut based method [16]. The former which determines the conclusion by geometrical transformation of the involved fuzzy sets. The latter executes the inference in three steps: 1. a state space transformation is performed, where the structure of the transformed space alleviate the possibility of abnormality; 2. the conclusion is calculated in the transformed space by the α-cut type inference method; 3. the conclusion is transformed back to the original state space. Finally this method tailors the conclusion as combination of KH methods for each characteristic point. We also remark that in [17] an axiomatic characterization of fuzzy interpolation is given.

B. Hierarchical reasoning

An interesting, though not well characterized, approach to reduce complexity is hierarchical reasoning, which is based on the modification of rule base structure, and thus reduce the exponent k in the expression of complexity. The main idea of this approach is that the multidimensional input state space $X = X_1 \times \cdots \times X_n$ can be decomposed, so that some of its components, e.g., $Z_0 = X_1 \times \cdots \times X_k$ determines a subspace ($k < n$), a so that in Z a partition $\Pi = \{D_1, \ldots, D_p\}$ can be determined. In each element of the partition Π, a sub-rule base can be constructed with local validity. In advantageous case, the local rule bases contain (much) less variables, and hence the complete structured rule base model has less complexity in terms of required time. The main problem of the method is that it is often difficult to determine a proper partition of the state space so, that in the elements of the partition some variables can be omitted. Despite this fact, the approach was successfully applied in the famous project of Sugeno's, where they used hierarchical fuzzy control of an unmanned helicopter [18].

There are attempts to combine hierarchical and sparse rule bases [19], which could lead to simultaneously reduce the base T and the exponent k in the expression of $O(T^k)$,

but the characterization of applicability and real-world applications based on such methods are still missing.

III. Reduction of data-based models

A. Orthogonal least squares-based techniques

The orthogonal least squares (OLS) method was first applied to fuzzy systems in [20], to select the most important fuzzy basis function needed to approximate a data set.

If we have N input-output data pairs $\{\mathbf{x}_j, y_j\}_{j=1}^{N}$ then (1) can be written as:

$$\mathbf{y} = \mathbf{W}b + \mathbf{e} \tag{2}$$

if the rules have only constant consequents b_i ($\mathbf{a}_i = 0$ for all the M rules), and where $\mathbf{W}$ is composed of vectors $w_i(\mathbf{x}_j)$ for $j \in [1, N]$, end $\mathbf{e}$ is the vector of approximation error.

The OLS method transforms the column of the firing matrix $\mathbf{W}$ into a set of orthogonal basis vectors in order to inspect the individual contribution of each rule. The Gram–Schmidt orthogonalization is used to decompose the matrix $\mathbf{W} = PA$, where $\mathbf{P}^T\mathbf{P} = I$ is an orthogonal matrix and A is an upper-triangular matrix. Substituting $\mathbf{W} = PA$ in Eq. (2) and using the fact that columns $\mathbf{p}_i$ of $\mathbf{P}$ are orthogonal, the sum of squares of $y(k)$ can be written as:

$$\mathbf{y}^T\mathbf{y} = \sum_{i=1}^{M} g_i \mathbf{p}_i^T \mathbf{p}_i + \mathbf{e}^T\mathbf{e}$$

where $g_i = A_i b_i$. The part of output variance $\mathbf{y}^T\mathbf{y}/N$ explained by regressors is $\sum g_i \mathbf{p}_i^T \mathbf{p}_i / N$. Thus an error reduction ration [21] due to an individual rule i can be determined:

$$[\mathrm{err}]^i = \frac{g_i \mathbf{p}_i^T \mathbf{p}_i}{\mathbf{y}^T\mathbf{y}}, quad 1 \leq i \leq M.$$

This ratio offers a simple means of ordering the rules, and was used in [20] to select a subset of important rules in a forward-regression manner.

However, as it was pointed out in [22], may produce an inappropriate subset of fuzzy rules. This problem was solved in [23], where the authors introduced a check for very small $\mathbf{p}_i^T\mathbf{p}_i$ value which is the effect of linear combination of column vectors $\mathbf{w}_i$ previously selected by the method.

The main advantage of OLS method is that it determines an importance ranking of the rules, so an the user can select the required number of rules based e.g. on the approximation accuracy or the relative contribution of the selected rules.

B. Singular value based decomposition

Higher Order SVD (HOSVD) based TS fuzzy approximation technique was initiated by *Yam* in 1997 [24], which directly finds the minimal number of rules from sampled values. Shortly after, this technique was introduced as SVD reduction of the rule base and structure decomposition by *Yam et al.* in [25]. An extension of *Yen* and *Wang's* work outlined above [22] to multi-dimensional (MIMO rule bases) cases may also be conducted in a similar fashion as the HOSVD reduction technique proposed in [24], [25]. These works give also some transformation techniques to yield antecedent sets with specific characteristics. Further extensions of [24] to a general rational form, inference algorithm independent rule bases, linguistic arrays and neural networks are proposed by *Baranyi et al.* in [26], [27], [28], [29].

The work in [27] can be applied regardless of the inference paradigm adopted for fuzzy rule base. Presumably, the product operation in [26] can be replaced by the Rudas's generalized inference operators [30], [31]. This would have a prominent role in developing the ability of finely tuning the TS models according to the application at hand and/or specific purposes of system performance. Work [32] specializes the use of SVD reduction to *Takagi-Sugeno* dynamic models. In order to facilitate further research a compact tensor product based notation is proposed by *Baranyi et al.* in [33], [34], [35], which works propose adaption techniques to HOSVD reduced rule bases. Various estimations for the error bound of HOSVD reduction techniques are proposed by *Takács* in [36], [37], [38], [39].

SVD is not merely used as a way of reduction of fuzzy rule bases. A brief enumeration of the potentials offered by SVD, of which some works were started by Beltarmi about 200 years ago, can be found in Stewart (1993) [40]. SVD is one of the most fruitful tools in linear algebra, belting its promising role in complexity reduction in general. The key idea of using SVD in complexity reduction is that the singular values can be applied to decompose a given system and indicate the degree of the significance of the decomposed parts. Reduction is conceptually obtained by the truncation of those parts which have weak or no contribution at all to the output according to the assigned singular values. This advantageous feature of SVD is used in *Yam's* works to extract a given model approximation and discard those rules which have no significant role in the overall system according to a given approximation accuracy. However, reducing the number of rules does not imply the computational cost reduction in all cases since the computation also depends on the number of overlapping antecedent membership functions, see [41], [42].

Let us briefly outline the key idea of HOSVD based reduction techniques. Let an N-variable rule base given with the following rules:

$$R_{i_1,i_2,\ldots,i_N}:$$

If x_1 is A_{1,i_1} and $\ldots x_N$ is A_{N,i_N} then y is $B_{i_1,i_2,\ldots,i_N}$.

Applying TS-model [9] the output is calculated as:

$$y = f(\mathbf{x}) = \sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\cdots\sum_{i_N=1}^{I_N}\prod_{n=1}^{N} w_{n,i_n}(x_n)y_{i_1,i_2,\ldots,i_N} \quad (3)$$

where vector $\mathbf{x} \in \Re^{\mathbf{N}}$ consists of the input values x_n, $n = 1\ldots N$ and $w_{n,i}(x)$ is the i-th antecedent function on the n-th input universe. Equation (3) can be described in tensor-product form as proposed in [33], [34]:

$$y = f(\mathbf{x}) = \mathbf{B}\bigotimes_{n=1}^{N}\mathbf{w}_n(x_n) \quad (4)$$

where tensor $\mathbf{B} \in \Re^{I_1\times I_2\times\ldots\times I_N}$ contains the consequent values $y_{i_1,i_2,\ldots,i_N}$. Row vector $\mathbf{w}(x_n)$ contains the membership functions of the antecedents as:

$$\mathbf{w}_n(x_n) = [w_{n,1}(x_n), w_{n,2}(x_n), .., w_{n,I_n}(x_n)] \quad (5)$$

Let briefly discuss the fundamentals of HOSVD in the sense of reduction. Many reduction properties of the HOSVD of higher-order tensors have been investigated in related literatures. Here, let us briefly summarize those that have prominent roles in complexity reduction and approximation. The HOSVD decomposes a given tensor $\mathbf{B} \in \Re^{I_1\times I_2\times\ldots\times I_N}$ into the product of:

$$\mathbf{B} = \mathbf{S}\bigotimes_{n=1}^{N}\mathbf{U}_n \quad (6)$$

(for notation see appendix) in which

1) matrix $\mathbf{U}_n$ is an unitary $(I_N \times I_N)$ matrix called n-mode singular matrix.

2) tensor $\mathbf{S} \in \Re^{I_1\times I_2\times\ldots\times I_N}$ of which the subtensors $\mathbf{S}_{i_n=\alpha}$ have the properties of

(i) all orthogonality: two tensors $\mathbf{S}_{i_n=\alpha}$ and $\mathbf{S}_{i_n=\beta}$ are orthogonal for all possible values of n, α and β: $\langle \mathbf{S}_{i_n=\beta}, \mathbf{S}_{i_n=\beta}\rangle = 0$ when $\alpha \neq \beta$.

(ii) ordering: $\|\mathbf{S}_{i_n=1}\| > \|\mathbf{S}_{i_n=2}\| > \ldots, \|\mathbf{S}_{i_n=I_n}\|$ for all possible values of n. The *Frobenius*-norm $\|\mathbf{S}_{i_n=i}\|$, symbolized by $\sigma_i^{(n)}$, are the n-mode singular values of $\mathbf{B}$.

First, in multi-linear algebra as well as in matrix algebra, the Frobenius-norm is unitary invariant. As a consequent, the squared Frobenius-norm of a matrix can be generalized as equal to the sum of its squared singular values. Let the approximation property of HOSVD be characterized first. Assume that tensor $\mathbf{B}$ is decomposed by HOSVD as above in eq. (6), further let the n-mode rank of $\mathbf{B}$ equals to R_n. Define tensor $\widehat{\mathbf{B}}$, by discarding singular values $\sigma_{I_n'+1}^{(n)}, \sigma_{I_n'+2}^{(n)}, \ldots, \sigma_{R_n}^{(n)}$ for given values of I_n', i.e. set the corresponding parts of $\mathbf{S}$ equal to zero. Then we have

$$\|\mathbf{B} - \widehat{\mathbf{B}}\| = \sum_{n=1}^{N}\sum_{i_n=I_n'+1}^{R_n}(\sigma_{i_n}^{(n)})^2 \quad (7)$$

This property is the higher-order equivalent of the link between the SVD of a matrix and its best approximation in a least-squares sense, by a matrix of lower rank. The situation is, however, quite different for tensors. By discarding the smallest n-mode singular values, one obtains a tensor $\widehat{\mathbf{B}}$ with n-mode rank of I_n' but this tensor is , in general, not the best possible approximation under the given n-mode rank constraints. Nevertheless, the ordering implies that the main components of $\mathbf{B}$ are mainly concentrated in the part corresponding to low values of the indices. Consequently, if $\sigma_{i_n}^{(n)} \gg \sigma_{i_n+1}^{(n)}$ where I_n' actually corresponds to the numerical ranking of $\mathbf{B}$ then the smaller n-mode singular values are not significant, justifying that they be discarded. In this sense, the $\widehat{\mathbf{B}}$ as obtained can still be considered as a good approximation of $\mathbf{B}$.

Consequently, if a rule base is given by tensor $\mathbf{B}$ and vectors $\mathbf{w}_n(x_n)$ as in eq. (4) its complexity can then be minimized via executing HOSVD on $\mathbf{B}$ as:

$$y = f(\mathbf{x}) = \mathbf{B}\bigotimes_{n=1}^{N}\mathbf{w}_n(x_n) = (\mathbf{S}\bigotimes_{n=1}^{N}\mathbf{U}_n)\bigotimes_{n=1}^{N}\mathbf{w}_n(x_n) = \quad (8)$$

$$= \mathbf{S}\bigotimes_{n=1}^{N}(\mathbf{w}_n(x_n)\mathbf{U}_n) \quad (9)$$

If singular values are discarded the size of $\mathbf{S}$ and $\mathbf{U}_n$ decreases. Let they respectively be denoted as $\mathbf{B}^r \in \Re^{I_1^r\times I_2^r\times\ldots\times I_N^r}$, where "$r$" means "reduced", and $\mathbf{U}_n^r \in \Re^{I_n\times I_n^r}$, where $I_n^r \leq I_n$ for all n. So, eq. (9) becomes:

$$y = f(\mathbf{x}) \cong \mathbf{B}^r\bigotimes_{n=1}^{N}(\mathbf{w}_n(x_n)\mathbf{U}_n^r) \quad (10)$$

finally we obtain:

$$y = f(\mathbf{x}) \cong \mathbf{B}^r\bigotimes_{n=1}^{N}\mathbf{w}_n^r(x_n) \quad (11)$$

where $\mathbf{w}_n^r(x_n) \in \Re^{I_n^r} = \mathbf{w}_n(x_n)\mathbf{U}_n^r$. eq. (11) represents a TS rule base like eq. (4). As a consequence the size of

$\mathbf{B}^r$ is less than $\mathbf{B}$ if singular values are discarded, hence the rule base is reduced.

An important issue should be addressed here. The reduced functions $w_{n,i}(x_n)$ in (11) may not be interpretable as membership functions, since the transformation using matrix $\mathbf{U}$ may result in functions with negative values. Another crucial point is that the resulted antecedent functions do not guarantee the *Ruspini*-partition, which means that the sum of the antecedents over all inputs may not be equal to 1. This fact would destroy the whole reduction concept. However, if only the saving of computational cost of final implementation is in purpose and no any conditions of the membership functions have to be accommodated then (11) is directly applicable. If the reduced form is for further studies in fuzzy theory then the reduced antecedent functions should accommodate additional characterization pertaining to specific operations. This may require further transformations. To obtain matrices $\mathbf{U}$ in such a way that the reduced antecedent functions are bounded by [0,1] and hold *Ruspini*-partition, *Non-Negativeness* (NN) and *Sum-Normalisation* (SN) transformation techniques are developed by *Yam* in [24], [25].

IV. Conclusions

In this paper we give a brief survey of reduction techniques. The motivation of this topic is that the identification of fuzzy models and controllers from training data needs to consider an important feature between data fitness and complexity. We emphasize the importance of these features by pointing out that fuzzy models and controllers with large number of local models may encounter the risk of having an approximation capable of fitting training data well, but incapable of running on low satisfactory computational cost. In order to help the developments of fuzzy models and controllers to strive for balance between the two conflicting design objectives, we survey various fuzzy model reduction techniques.

V. Appendix

DEFINITION (**n-mode matrix of tensor A**) *Assume N-th order tensor* $\mathbf{A} \in \Re^{I_1 \times I_2 \times \ldots \times I_N}$. *The n-mode matrix* $\mathbf{A}_{(n)} \in \Re^{I_n \times J}$, $J = \prod_k^{I_n}$, *where* $k = 1 \ldots N$ *and* $k \neq n$, *contains all the vectors of the n-th dimension of tensor* $\mathbf{A}$. *The ordering of the vectors is arbitrary, this ordering shall, however be consistently used later on.*

DEFINITION (**n-mode matrix-tensor product**) *The n-mode product of tensor* $\mathbf{A} \in \Re^{I_1 \times I_2 \times \ldots \times I_N}$ *by a matrix* $\mathbf{U} \in \Re^{J \times I_n}$ *denoted by* $\mathbf{A} \times_n \mathbf{U}$ *is an* $(I_1 \times I_2 \times \ldots \times J \times \ldots \times I_N)$*-tensor of which the entries are given by* $\mathbf{A} \times_n \mathbf{U} = \mathbf{B}$, *where* $\mathbf{B}_{(n)} = \mathbf{U}\mathbf{A}_{(n)}$. *Let* $\mathbf{A} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \ldots \times_N \mathbf{U}_N$ *be noted for brevity as* $\mathbf{A}^r \bigotimes_{n=1}^{N} \mathbf{U}_n$.

DEFINITION (**n-mode rank of tensor A**) *The n-mode rank of tensor* $\mathbf{A}$ *is the rank of the n-mode matrix of tensor* $\mathbf{A}$: $rank_n(\mathbf{A}) = rank(\mathbf{A}_{(n)})$.

References

[1] L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Trans. on SMC*, vol. 3, pp. 28–44, 1973.

[2] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *Int. J. of Man Machine Studies*, vol. 7, pp. 1–13, 1975.

[3] L. X. Wang, "Fuzzy systems are universal approximators," in *Proc. of the IEEE Int. Conf. on Fuzzy Systems*, San Diego, 1992, pp. 1163–1169.

[4] J. J. Buckley, "Sugeno type controllers are universal controllers," *Fuzzy Sets and Systems*, vol. 53, pp. 299–304, 1993.

[5] J. L. Castro, "Fuzzy logic controllers are universal approximators," *IEEE Trans. on SMC*, vol. 25, pp. 629–635, 1995.

[6] E. P. Klement, L. T. Kóczy, and B. Moser, "Are fuzzy systems universal approximators?," *Int. J. General Systems*, vol. 28, no. 2–3, pp. 259–282, 1999.

[7] B. Moser, "Sugeno controllers with a bounded number of rules are nowhere dense," *Fuzzy Sets and Systems*, vol. 104, no. 2, pp. 269–277, 1999.

[8] D. Tikk, "On nowhere denseness of certain fuzzy controllers containing prerestricted number of rules," *Tatra Mountains Math. Publ.*, vol. 16, pp. 369–377, 1999.

[9] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. on SMC*, vol. 15, pp. 116–132, 1985.

[10] M. Sugeno and G. T. Kang, "Structure identification of fuzzy model," *Fuzzy Sets and Systems*, vol. 28, pp. 15–33, 1988.

[11] L. T. Kóczy, "Computational complexity of various fuzzy inference algorithms," *Annales Univ. Sci. Budapest, Sect. Comp.*, vol. 12, pp. 151–158, 1991.

[12] L. T. Kóczy and K. Hirota, "Approximate reasoning by linear rule interpolation and general approximation," *Internat. J. Approx. Reason.*, vol. 9, pp. 197–225, 1993.

[13] L. T. Kóczy and K. Hirota, "Size reduction by interpolation in fuzzy rule bases," *IEEE Trans. on SMC*, vol. 27, pp. 14–25, 1997.

[14] D. Dubois and H. Prade, "Gradual rules in approximate reasoning," *Information Science*, vol. 61, pp. 103–122, 1992.

[15] P. Baranyi and L. T. Kóczy, "A general and specialized solid cutting method for fuzzy rule interpolation," *BUSEFAL*, vol. 67, pp. 13–22, 1996.

[16] D. Tikk and P. Baranyi, "Comprehensive analysis of a new fuzzy rule interpolation method," *IEEE Trans. on Fuzzy Systems*, vol. 8, no. 3, pp. 281–296, 2000.

[17] S. Jenei, "A new approach for interpolation and extrapolation of compact fuzzy quantities," in *Proc. of the 21th Linz Seminar on Fuzzy Set Theory*, 2000, pp. 13–18.

[18] M. Sugeno, M. F. Griffin, and A. Bastian, "Fuzzy hierarchical control of an unmanned helicopter," in *Proc. of the 5th IFSA World Congress (IFSA'93)*, Seoul, 1993, pp. 1262–1265.

[19] L. T. Kóczy, K. Hirota, and Muresan, "Interpolation in hierarchical fuzzy rule bases," *International Journal of Fuzzy Systems*, vol. 1, no. 2, 1999.

[20] L. X. Wang and J. M. Mendel, "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning," *IEEE Trans. on Neural Networks*, vol. 3, pp. 807–813, Oct. 1992.

[21] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. on Neural Networks*, vol. 2, pp. 302–309, Apr. 1991.

[22] J. Yen and L. Wang, "Simplifying fuzzy rule-based models using orthogonal transform methods," *IEEE Trans. on Syst., Man, Cybern. B*, vol. 29, pp. 13–24, Feb. 1999.

[23] M. Setnes and R. Babuška, "Rule base reduction: some comments on the use of orthogonal transforms," *IEEE Trans. on Syst., Man, Cybern. C*, vol. 31, no. 2, pp. 199–206, May 2001.

[24] Y. Yam, "Fuzzy approximation via grid point sampling and singular value decomposition," *IEEE Trans. SMC*, vol. 27, pp. 933–951, 1997.

[25] Y. Yam, P. Baranyi, and C. T. Yang, "Reduction of fuzzy rule base via singular value decomposition," *IEEE Transaction on Fuzzy Systems*, vol. 7, pp. 120–131, 1999.

[26] P. Baranyi and Y. Yam, "Fuzzy rule base reduction," *Chapter 7 of Fuzzy IF-THEN Rules in Computational Intelligence: Theory and Applications Eds., D. Ruan and E.E. Kerre, Kluwer*, vol. 7, pp. 135–160, 2000.

[27] P. Baranyi, Y. Yam, C. T. Yang, P. Varlaki, and P. Michelberger, "Inference algorithm independent svd fuzzy rule base complexity reduction," *International Journal of Advanced Computational Intelligence*, vol. 5, pp. 22.

[28] P. Baranyi, Y. Yam, P. Varlaki, and P. Michelberger, "Singular value decomposition of linguistically defined relations," *International Journal of Fuzzy Systems*, vol. 2, pp. 108.

[29] P. Baranyi K.Lei and Y. Yam, "Complexity minimalisation of non-singleton based fuzzy-neural network," *International Journal of Advanced Computational Intelligence*, vol. 4, pp. 1.

[30] I.J. Rudas and O.Kaynak, "New types of generalized operations," *Computational Intelligence: Soft Computing and Fuzzy-Neuro Integrationwith Applications. Springer NATO ASI Series. Series F: Computer and Systems, Sciences,*, vol. 192, pp. 128–156, 1998.

[31] I.J. Rudas, "Towards the generalization of t-operators: a distance-based approach," *Jour. of Information and Organizational Sciences.*, vol. 23, pp. 149–166, 1999.

[32] P. Baranyi, Y. Yam, A. Vrkonyi-Kczy, R. J. Patton, P. Michelberger, and M. Sugiyama, "Svd based reduction to ts fuzzy models," *IEEE Transaction on Industrial Electronics.*

[33] P. Baranyi, Y. Yam, D. Tikk, and R. J. Patton, *Trade-off between approximation accuracy and complexity: TS controller design via HOSVD based complexity minimization*, Studies in Fuzziness and Soft Computing. Physica-Verlag, Heidelberg, trade-off between accuracy and interpretability in fuzzy rule-based modelling edition, 2002.

[34] P. Baranyi, A. R. Varkonyi-Koczy, Y. Yam, and P. Michelberger, "Hosvd based computational complexity reduction of ts fuzzy models," *IEEE 9th IFSA World Congress.*

[35] P. Baranyi, A. R. Varkonyi-Koczy, Y. Yam, P. Varlaki, and P. Michelberger, "An adaption technique to svd reduced rule bases," *IEEE 9th IFSA World Congress.*

[36] O. Takacs and A. R. Varkonyi-Koczy, "Error-bound for the non-exact svd-based complexity reduction of the generalized hybrid neural networks with non-singleton consequents," *IEEE Instrumentation and Measurement Technology Conference.*

[37] O. Takacs and A. R. Varkonyi-Koczy, "Non-exact complexity-reduction of generalized fuzzy-neural networks," *MFT Periodika 2001-02, www.mft.hu.*

[38] O. Takacs and A. R. Varkonyi-Koczy, "Svd-based complexity reduction of rule-bases with nonlinear antecedent fuzzy sets," *IEEE Int. Workshop on Intelligent Signal Processing.*

[39] O. Takacs and I. Nagy, "Error bound of the svd based neural networks," *IFAC Symposium, Artificial Intelligence in Real Time Control.*

[40] G. W. Stewart, "On the early history of the singular value decomposition," *SIAM Rev.*, vol. 35, pp. 551.

[41] Y. Yam, P. Baranyi, and C. T. Yang, "Author's replay," *IEEE Transaction on Fuzzy Systems*, vol. 9, pp. 676–677, 2001.

[42] C. W. Tao, "Comments on reduction of fuzzy eule base via singular value decomposition," *IEEE Transaction on Fuzzy Systems*, vol. 9, pp. 675–676, 2001.