Do Hierarchical Fuzzy Systems Really Improve Interpretability?

Luis Magdalena(✉)

Escuela Técnica Superior de Ingenieros Informáticos,
Universidad Politécnica de Madrid, Campus de Montegancedo,
28660 Boadilla del Monte, Madrid, Spain
luis.magdalena@upm.es

Abstract. Fuzzy systems have demonstrated a strong modeling capability. The quality of a fuzzy model is usually measured in terms of its accuracy and interpretability. While the way to measure accuracy is in most cases clear, measuring interpretability is still an open question.

The use of hierarchical structures in fuzzy modeling as a way to reduce complexity in systems with many input variables has also shown good results. This complexity reduction is usually considered as a way to improve interpretability, but the real effect of the hierarchy on interpretability has not really been analyzed.

The present paper analyzes that complexity reduction comparing it with that of other techniques such as feature extraction, to conclude that only the use of intermediate variables with meaning (from the point of view of model interpretation) will ensure a real interpretability improvement due to the hierarchical structure.

Keywords: Fuzzy · Hierarchical · Interpretability · Semantics
Complexity

1 Introduction

A model is the representation of a system (a part of the world). As a consequence, modeling is the process of creating a representation (model) of a certain system. The model can take quite different forms ranging from physical (a mockup) to formal models. Formal models use rules, concepts, mathematical equations, etc. to describe the system; and represent a powerful analysis tool.

As the model is a representation of the system, evaluating its quality usually encompasses different aspects that strongly relate to the purpose of the model. If the model was simply built as a demonstration tool, to show a client how the final system will look, it will only need to capture the essence, the idea of the real system. Other models are designed to know how the system will behave in the presence of a certain stimulus, as the model of an airplane wing to be tested in a wind tunnel. That kind of situation requires the model behaving as close as possible to the real system.

J. Medina et al. (Eds.): IPMU 2018, CCIS 853, pp. 16–26, 2018.
https://doi.org/10.1007/978-3-319-91473-2_2

When analyzing a computer model to evaluate its quality, it is also possible to consider different aspects. If the purpose of the model is similar to that of the airplane wing model tested on a wind tunnel, the idea will be to know how will the wing behave under certain wind conditions. Consequently, the closer the behavior of model and system was, the better the model will be. When that is the situation, it can be properly managed with a formal model that *simply* replicates the input-output relations of the system, with no particular interest on how it does. This task is well suited for many modeling tools including those known as black-box models. A completely different situation is that of a modeling process in which we are interested not only in *what* will be the output, but also in *why* will it be such. It is clear that the pure input-output relation is not enough in the latter case. The presentation of pieces of knowledge describing or explaining that input-output relation is needed, and consequently the internal structure of the model will be capital to cope with this kind of situations.

Summarizing the previous ideas, we can say that the quality of a model can be measured in terms of **how accurately reproduces** the stimulus/response relation of the modeled system, but also in terms of **how clearly it explains** or describes the underlying mechanism producing, or the knowledge justifying, those input-output relations.

Among the many tools that have been used for modeling, fuzzy systems have demonstrated great performance when applied to many real world problems. System modeling with fuzzy rule-based systems (FRBSs) is usually known as fuzzy modeling (FM) [3]. A fuzzy model, as any other model, can be evaluated in terms of those two previously described concepts: how accurately reproduces the behavior, and how clearly describes the underlying knowledge. Fuzzy models are well suited for both questions: the *accuracy*, capability to faithfully represent the real system, and the *interpretability*, capability to express the behavior of the real system in an understandable way. But when both of them are jointly considered, they mostly appear as two contradictory requirements. In fact, literature initially established subareas focusing on one or the other. While linguistic FM (mainly developed by linguistic FRBSs) was focused on the interpretability, precise FM (mainly developed by Takagi-Sugeno-Kang FRBSs) was focused on the accuracy. At present, both criteria are considered of vital importance, so that the balance between them has gained a significant attention in the field [5,6].

While accuracy can easily be measured (e.g., in terms of errors), interpretability evaluation still represents an open question where many different concepts and metrics offer a wide repertory of options. There is at least a certain level of agreement in considering the existence of two types of interpretability [8]: related to complexity and related to semantics. Semantic based metrics [2,9,17] have recently appeared to complement or complete the preexisting complexity based metrics [18]. Different overviews and comparisons of interpretability approaches can be considered [8,10], but only recently, the question of interpretability has been considered in the framework of type-2 FRBS [1,14,15] or hierarchical fuzzy systems (HFS) [19,23].

Taking into account that the primal idea for introducing hierarchical fuzzy systems was related to the reduction of structural complexity, namely, to avoid

the course of dimensionality appearing in conventional FRBSs, it is clear that the interaction between interpretability and hierarchical fuzzy systems is a question to be considered. Nevertheless, only a few authors have studied it, as previously said. The present paper will focus on this question by analyzing the relation between hierarchy and interpretability, concentrating on the semantic component of interpretability.

2 Hierarchical Fuzzy Systems

When designing a FRBS to model a complex problem (particularly those with a large number of input variables) designers must cope with what is usually known as the *curse of dimensionality*, i.e., the exponential growth in the number of rules related to the number of input variables. Different options have been considered to manage this challenging questions: the use of compact rule structures, sparse rule bases, or a hierarchical fuzzy system (HFS) among others.

The way to create the hierarchical structure is not unique, and it is possible to define different kinds of hierarchy in the fuzzy system (different structures). The main difference relates to the components of the overall system being affected by the hierarchical decomposition. Three main options can be described: decompose at the level of fuzzy partitions, at the level of variables, or at the level of rules.

A *hierarchy of rules* produces a prioritization in the use of the rules in such a way that more specific rules receive a higher priority, while priority is lower for more generic rules [21,22]. In this approach a generic rule is applied only when no applicable specific rule is available, and the rules are grouped into prioritized levels to design an HFS. This structure has clear effects from the point of view of output explanation, where interpreting the output involves using the concept of *level of specificity* of the rule.

Other authors establish a *hierarchy of partitions* for each variable, with different levels of granularity. With this concept, the hierarchical structure is composed of a set of layers where each one contains linguistic partitions (concerning all the same set of variables) with different granularity levels, and linguistic rules whose linguistic variables take values in these partitions. The idea is clearly related to that of generic/specific rules, where the specificity of the rule relies on the specificity of the partition. The main difference concerns the design process that in this case is systematic, based either in reduction [11,13] or expansion [7] methods.

But the most common approach to HFSs, and the one we will consider in this paper, is that of the *hierarchy of variables*. The idea for these HFSs is to split a large system into a cascade of several smaller systems, by decomposing the input space into several input spaces with a reduced number of variables, where each input variable is only considered at a certain level of the hierarchy (Fig. 1). To involve all variables in the generation of the overall output, the output of each level is considered as one of the inputs to the following level [12].

It is clear that the main effect achieved with this hierarchical process is the reduction of the number of rules of the FRBS, i.e., the palliation of the curse

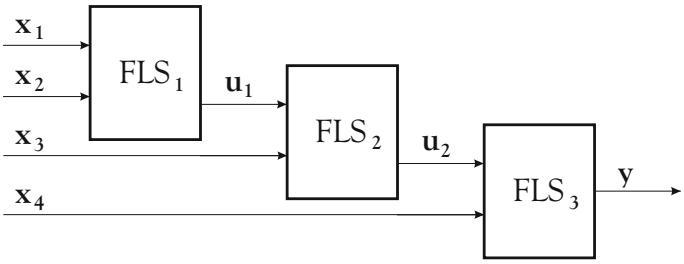

Fig. 1. Hierarchical fuzzy system (serial)

of dimensionality problem. As an example, a system with n input variables and m linguistic labels per variable will have m^n rules in a conventional FRBS. Transformed in a hierarchical fuzzy controller where the n variables are divided into L different levels, with n_k variables (including the output variable of the previous level) as inputs to the k^{th} level of the hierarchy, the total number of rules is given by

$$T = \sum_{k=1}^{L} m^{n_k} \tag{1}$$

with

$$n_1 + \sum_{k=2}^{L} (n_k - 1) = n \tag{2}$$

And this number of rules will take on its minimum value when $n_k = 2$ (Figs. 1 and 2), being this minimum equal to

$$T = (n-1) * m^2$$

In summary, the number of rules in a complete hierarchical rule base could be reduced to a linear function of the number of variables, while in a conventional FRBS it was an exponential function of the number of variables.

In addition to the hierarchical structure shown in Fig. 1, usually known as incremental or serial HFS, where each level contains a single Fuzzy Systems, it is possible to define other hierarchical structures. The so called parallel or aggregated HFS receives all input variables in the fuzzy systems located at first level, having only output variables from the previous level as inputs of the subsequent level (Fig. 2).

Finally, cascade fuzzy systems [16,20] represent another option where all input variables are considered at every level of the hierarchy (in addition to previous levels outputs) somehow loosing the potential to cope with dimensionality problems.

Quite recently, the concept of cascade fuzzy systems have been revisited [23] to define the stacked hierarchical structure to improve its interpretability by additional complexity reduction. But this approach does not focus on what will be discussed below.

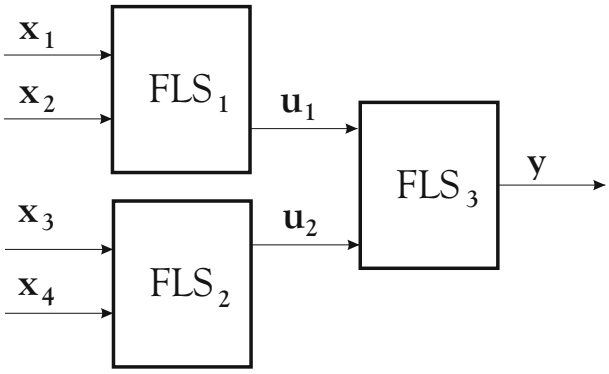

Fig. 2. Hierarchical fuzzy system (parallel)

3 Interpretability in HFSs

As interpretability has been widely linked to complexity (the higher the complexity, the lower the interpretability), the complexity reduction provided by HFSs has been usually viewed as the proof of interpretability improvement produced by these systems. However, the attempts to directly analyze HFSs in terms on interpretability (without putting it behind complexity) started only quite recently. What probably is the first approach to interpretability analysis for parallel and serial HFSs is presented in [19].

The idea of that paper is to use conventional interpretability measures to evaluate interpretability at the level of each of the FLSs composing the HFS (three in Figs. 1 and 2), and then aggregate the obtained values by means of a weighted sum that finally produces a value between 0 and 1. The aggregation works first with FLSs in the same layer to which the average is applied. Once obtained a single value per layer, different weights are applied to each layer, being higher for layers closer to the input and lower for those closer to the output (so descending when advancing through the hierarchy). The rationale behind this structure of weights is that usually the most influential variables are considered at first layer, and each new layer applies the most influential of the remaining ones, so that the output layer considers the least influential variables. In that way, the layers that apply more influential variables have a higher contribution to overall interpretability than those using less influential variables.

The main idea underlying the approach is that a hierarchical structure allows the independent analysis of the different blocks building up the hierarchy. The subsystems are considered as decoupled structures that can be independently interpreted. But the question is: Is it true? Is it really possible to interpret each subsystem as a single entity? We think that decoupling the analysis is only possible under certain circumstances, as will be considered below.

3.1 Structuring the Variables to Reduce Complexity

Hierarchical approaches are not the only way to cope with complexity in FRBSs. Many other options are possible and have been widely considered in literature. But if we focus on the idea of a hierarchy of variables, i.e., structuring the variables in different levels, it seems that the closest approaches (from a conceptual point of view) are those centered on feature extraction and selection. Complexity reduction through feature extraction and feature selection has a large presence not only in fuzzy systems but in almost any modeling technique. But there is a significant difference between feature extraction and feature selection. While feature extraction creates new *synthetic variables* encompassing the information proceeding from several variables, feature selection does not create any new variable, it simply picks up a few of the preexisting variables, those that apparently better represent the overall system.

Feature extraction generates a reduced set of new features from the original set, by means of a mapping function, trying to represent the original data more concisely. But this process is computationally expensive and, what is more interesting in this scope, produces a loss of interpretability since in most cases (probably always) no explicit and intuitive (semantic) relation exists between the original and the new features, being the original features the only ones having a *physical* explanation.

This point is somehow implicitly accepted by any designer of *interpretable fuzzy systems*, but, if we consider the many different interpretability metrics available in literature, to the author's knowledge no one will support this idea. If we generate a model with the same number of variables, rules, linguistic labels, etc, no index will consider how meaningful were the input variables, and consequently no one will distinguish between a model using selected variables and a second one using extracted (meaningless) variables.

It can be argued that the different measures and criteria are designed to compare several models designed in a similar context, or under similar boundary conditions, i.e., if variables are meaningless, that is something that can not be solved and will similarly affect any possible model. In that sense, we want to obtain the *best possible* model assuming the starting point. Consequently, feature extraction/selection is considered as a preliminary step were we can decide to avoid the meaningless variables.

Apparently, the approach better suited to perceive the differences between selected and extracted variables will be the *logical view index* based on cointension [4,17]. In this approach, cointension refers to a relation between concepts such that two concepts are cointensive if they refer to the same objects. Thus, a knowledge base will be interpretable if its semantics is cointensive with the knowledge a user builds in his/her mind after reading the knowledge representation (expressed in natural language). And we can consider that synthetic (extracted) variables will not be cointensive with any knowledge *in the mind of the reader*. In any case, there is not a clear way to measure how meaningful/meaningless are the variables. In addition, the implementation of this method focus on internal aspects of the designed model and not in the selection of input

variables. Further exploration of cointension as a way to evaluate semantic quality of synthetic variables could be an option to cope with this question.

3.2 The Need for Using Meaningful Intermediate Variables

When measuring the interpretability of a linguistic variable, existing metrics only pay attention to questions as number of terms, distinguishability of fuzzy sets, coverage of the universe of discourse, etc. Properties that do not rely on the conceptual interpretation of the variable. Consequently there is no formal way to assert that a fuzzy system with selected features is more interpretable than the one with extracted variables, since no interpretability measure will make any difference. In fact, this conceptual interpretability seems to be a rather subjective question, and consequently, almost impossible to capture with the kind of *objective* metrics used to measure interpretability. But it is commonly accepted that selected variables are more interpretable than extracted features.

And what is the relation between the feature selection/extraction question, and hierarchical fuzzy systems design? The parallelism is quite simple, intermediate variables in hierarchical fuzzy systems are, at the end, equivalent to extracted features. The only difference is the kind of functional relations between original and extracted variables (mostly arithmetical) or input and intermediate variables (defined in terms of fuzzy rules). In that way, the fuzzy system with feature extraction will somehow be equivalent to a parallel two-levels hierarchical fuzzy system where the first level comprises the synthesis of the extracted variables, and the second level is made up of the fuzzy system itself (Fig. 3). So, why do we assume that the use of extracted variables reduces interpretability while the use of hierarchical systems increases it?

The most plausible answer is that we assume intermediate variables in a hierarchical fuzzy systems are *meaningful* as they are generated as the output of an *interpretable fuzzy system*. But most HFSs use intermediate variables without any conceptual/semantic support, i.e., synthetic variables.

Let us assume we have a four inputs system with three terms per partition. According to previous analysis, the minimal hierarchical structure will only need $3 \times 3^2 = 27$ rules, while a conventional system will contain $3^4 = 81$ rules. It should be much easier to interpret 27 rules with two variables per rule than 81 rules with four variables per rule. But in the first case we are assuming that each subsystem (FLS_n) can be interpreted by itself. And how can by independently interpreted a system where one (several) of the variables *has no meaning*? We can imagine a rule like

> When *First variable* is *Low* and *Second variable* is *Medium* then *Temperature* is *High*

It looks impossible to interpret this rule alone, without knowing what do *First* and *Second variable* mean. So, we need to analyze the system as a whole, including the definition of the intermediate variables (First and Second). The possibility of decoupling the analysis of a complex system into several simpler systems

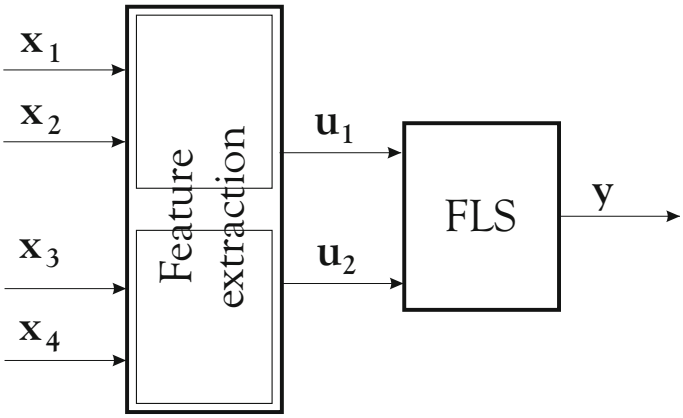

Fig. 3. Feature extraction as a hierarchical structure

seems to be lost. The different blocks are so tightly coupled that the only option is analyzing them as a whole. Let us then consider the complexity of the system as a whole.

For simplicity we will assume that the considered structure is that of Fig. 2, but the results would be identical with a serial structure. Intermediate variables will have three associated linguistic terms (the same for input variables) being $\{B_{11}, B_{12}, B_{13}\}$ the term set for u_1 and $\{B_{21}, B_{22}, B_{23}\}$ the term set for u_2. The fuzzy systems at first hierarchical level (FLS_1 and FLS_2) will contain nine rules per system, each of those rules having a consequent of the form u_i is B_{ij} with $i = 1, 2$ and $j = 1, 2, 3$. Let us define n_{ij} as the number of rules in FLS_i that refer as output to term B_{ij}. It is clear that $\sum_{j=1}^{3} n_{ij} = 9$ (the number of rules in the subsystem).

If we need to consider the hierarchical system as a single block, each rule in the second level of the hierarchy should be connected to the corresponding rules in first level, to be interpreted. And how should we connect them? We must expand each second level rule with all first level rules activating it. Consequently the rule

$$\text{IF } u_1 \text{ is } B_{11} \text{ and } u_2 \text{ is } B_{21} \text{ then } y \text{ is } C_k,$$

should be expanded with every rule from FLS_1 having B_{11} as output, and with every rule from FLS_2 having B_{21} as output. And the result is that the first rule in FLS_3 will be expanded to $n_{11} \times n_{21}$ rules considering four input variables each. If we repeat the process with the nine rules in FLS_3, the result is that the overall number of rules will be

$$\sum_{i=1}^{3}\sum_{j=1}^{3} n_i \times n_j = \sum_{i=1}^{3} n_i \times \sum_{j=1}^{3} n_j = 9 \times 9 = 3^4,$$

i.e., the same number of rules than the original (non hierarchical) fuzzy system. Then, the conclusion is that a hierarchical fuzzy system where the intermediate

variables do not allow us to decouple the interpretation into subsystems, does not produce a real reduction of complexity from the point of view of interpretability, and consequently does not improve structural interpretability.

In summary, the use of a hierarchical structure only improves interpretability when provides us with the capability of decoupling the overall system into a family of simpler subsystems that can be interpreted independently. And this is only possible when intermediate variables are meaningful from the point of view of interpretation. Otherwise, using a hierarchical structure is not significantly different than the use of feature extraction techniques. The use of synthetic features as intermediate variables, regardless how those features were created (either with fuzzy rules or with a function), avoids decoupling and consequently makes impossible a proper interpretation of subsystems as independent entities.

4 Conclusions

The only way to ensure an actual improvement of interpretability in an HFS is by means of a semantic-guided design of the hierarchy, where the selected blocks of variables produce subsystems with independent meaning characterized by the appearance of intermediate variables linked to properties of the represented system, i.e., intermediate variables with meaning. Any *blind* approach synthesizing intermediate variables without any semantic relation to the modeled system, simply hides the real complexity of the system. From the interpretation point of view, an HFS using meaningless variables maintains the same complexity (number of variables, rules, terms, etc) than the non-hierarchical one.

When considering a hierarchical fuzzy system to analyze its interpretability, there is a key question:

> Is it possible to independently interpret each of the multiple fuzzy systems building up the hierarchy?

And there are only two options. If we can interpret each FLS in the hierarchy as a single entity with their inputs and outputs, and understand the role of that piece of knowledge in the overall system, the hierarchy is really improving interpretability by means of an actual complexity reduction. If the FLSs are not interpretable alone, mainly due to the fact that their input and output variables are not linked to the problem under consideration, the hierarchy does not really improve interpretability. And the way to connect those variables to the problem is by linking the intermediate variables added when building the hierarchical structure, to properties, features, characteristics, etc., of the modeled system.

In early times of fuzzy modeling, some designers considered that *any* fuzzy system was interpretable. Later on it was commonly accepted that interpretability was not an intrinsic property of fuzzy models, but something achieved through a suitable structure and design process. Now this same idea should be extended to HFSs. They are not intrinsically more interpretable than *conventional* fuzzy systems. Its interpretability relates, at least, to the appropriate selection of intermediate variables. Further analysis, as well as the definition of metrics adapted to measure interpretability of HFSs will be the matter for future works.

Acknowledgements. This paper was partially supported by Universidad Politécnica de Madrid (Spain).

References

1. Alhaddad, M., Mohammed, A., Kamel, M., Hagras, H.: A genetic interval type-2 fuzzy logic-based approach for generating interpretable linguistic models for the brain P300 phenomena recorded via brain–computer interfaces. Soft. Comput. **19**(4), 1019–1035 (2015)
2. Alonso, J.M., Magdalena, L., Guillaume, S.: HILK: a new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. Int. J. Intell. Syst. **23**(7), 761–794 (2008)
3. Babuska, R.: Fuzzy Modeling and Control. Kluwer, Norwell (1998)
4. Cannone, R., Alonso, J.M., Magdalena, L.: Multi-objective design of highly interpretable fuzzy rule-based classifiers with semantic cointension. In: IEEE Symposium Series on Computational Intelligence (IEEE-SSCI), IV International Workshop on Genetic and Evolutionary Fuzzy Systems (GEFS), Paris, pp. 1–8 (2011)
5. Casillas, J., Cordon, O., Herrera, F., Magdalena, L.: Interpretability Issues in Fuzzy Modeling. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-37057-4
6. Casillas, J., Cordón, O., Herrera, F., Magdalena, L. (eds.): Accuracy Improvements in Linguistic Fuzzy Modeling. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-37058-1
7. Cordón, O., Herrera, F., Zwir, I.: Linguistic modeling by hierarchical systems of linguistic rules. IEEE Trans. Fuzzy Syst. **10**(1), 2–20 (2002)
8. Gacto, M.J., Alcala, R., Herrera, F.: Interpretability of linguistic fuzzy rule-based systems: an overview of interpretability measures. Inf. Sci. **181**(20), 4340–4360 (2011)
9. Galende, M., Gacto, M., Sainz, G., Alcalá, R.: Comparison and design of interpretable linguistic vs. scatter FRBSs: Gm3m generalization and new rule meaning index for global assessment and local pseudo-linguistic representation. Inf. Sci. **282**, 190–213 (2014)
10. Guillaume, S.: Designing fuzzy inference systems from data: an interpretability-oriented review. IEEE Trans. Fuzzy Syst. **9**(3), 426–443 (2001)
11. Guillaume, S., Charnomordic, B.: Generating an interpretable family of fuzzy partitions from data. IEEE Trans. Fuzzy Syst. **12**(3), 324–335 (2004)
12. Raju, G.V.S., Zhou, J., Kisner, R.A.: Hierarchical fuzzy control. Int. J. Control **54**(5), 1201–1216 (1991)
13. Ishibuchi, H., Nozaki, K., Yamamoto, N., Tanaka, H.: Selecting fuzzy if-then rules for classification problems using genetic algorithms. IEEE Trans. Fuzzy Syst. **3**(3), 260–270 (1995)
14. Juang, C.F., Chen, C.Y.: Data-driven interval type-2 neural fuzzy system with high learning accuracy and improved model interpretability. IEEE Trans. Cybern. **43**(6), 1781–1795 (2013)
15. Lucas, L., Centeno, T., Delgado, M.: Towards interpretable general type-2 fuzzy classifiers. In: 9th International Conference on Intelligent Systems Design and Applications, ISDA 2009, pp. 584–589 (2009)
16. Mar, J., Lin, H.T.: A car-following collision prevention control device based on the cascaded fuzzy inference system. Fuzzy Sets Syst. **150**(3), 457–473 (2005)

17. Mencar, C., Castiello, C., Cannone, R., Fanelli, A.: Interpretability assessment of fuzzy knowledge bases: a cointension based approach. Int. J. Approx. Reason. **52**(4), 501–518 (2011)
18. Nauck, D.: Measuring interpretability in rule-based classification systems. In: Proceedings of 12th IEEE International Conference on Fuzzy Systems, vol. 1, pp. 196–201. IEEE (2003)
19. Razak, T., Garibaldi, J., Wagner, C., Pourabdollah, A., Soria, D.: Interpretability indices for hierarchical fuzzy systems. In: IEEE International Conference on Fuzzy Systems. Institute of Electrical and Electronics Engineers Inc. (2017)
20. Wang, S., Chung, F., HongBin, S., Dewen, H.: Cascaded centralized tsk fuzzy system: universal approximator and high interpretation. Appl. Soft Comput. J. **5**(2), 131–145 (2005)
21. Yager, R.R.: On a hierarchical structure for fuzzy modeling and control. IEEE Trans. Syst. Man Cybern. **23**(4), 1189–1197 (1993)
22. Yager, R.R.: On the construction of hierarchical fuzzy systems models. IEEE Trans. Syst. Man Cybern. **28**(1), 55–66 (1998)
23. Zhang, Y., Ishibuchi, H., Wang, S.: Deep takagi-sugeno-kang fuzzy classifier with shared linguistic fuzzy rules. IEEE Trans. Fuzzy Syst. (in Press)