HFRBC-GA: A fuzzy classifier for energy systems applications

Eleonora D'Andrea and Beatrice Lazzerini

Dipartimento di Ingegneria dell'Informazione
University of Pisa
Largo Lucio Lazzarino 1, 56122 Pisa, Italy
eleonora.dandrea@for.unipi.it, b.lazzerini@iet.unipi.it

Abstract—**This paper presents a hierarchical approach for building fuzzy classifiers directly from data following a multi-level grid-like partition of the input domain. The fuzzy classifier is actually the union of several fuzzy systems built on input domain regions increasingly smaller. In order to guarantee high interpretability and to avoid the explosion of the number of rules, only the necessary partitions are built as the hierarchical level increases. Finally, a genetic algorithm is employed to optimize some free parameters of the proposed methodology. The method has been validated on 10 well-known benchmark datasets, by showing how the achieved results compare favorably with those obtained by other fuzzy classifiers in the literature. In addition, we apply our method to three case studies related to energy systems. In the first case study we linguistically describe how the solar irradiation and the temperature of the photovoltaic (PV) panel relate to the quantity of energy produced by a PV installation. The second and third case studies refer to the estimation of energy consumption in buildings. More precisely, we describe how the solar irradiation affects the use of electric lighting, and how the outdoor temperature impacts on hot water boiler usage.**

Keywords-energy consumption; fuzzy rule-based classifier; genetic algorithm; hierarchical approach; pattern classification; photovoltaic energy

I. INTRODUCTION

Fuzzy rule-based systems have been widely applied to pattern classification problems thanks to their capability to achieve good trade-offs between accuracy and interpretability [1, 2]. In particular, interpretability of a fuzzy rule-based system is typically measured in terms of complexity of the rule base, and depends on such factors as comprehensibility of fuzzy partitions of the domains of the involved linguistic variables, number of input variables, number of conditions in the antecedent of each rule, and number of fuzzy rules. In its simplest form, a fuzzy rule-based classifier (FRBC) is a system consisting of fuzzy if-then rules having a class label as consequent.

When designing an FRBC two main problems must be considered: fuzzy classifier identification and fuzzy parameter optimization. Further, regarding the former, major issues are i) how to choose the membership functions of linguistic variables, ii) how to generate the fuzzy rules, and iii) how to determine the output class.

A large number of methods for extracting fuzzy rules directly from numerical data have been proposed, thus making prior knowledge about the data unnecessary. These methods include heuristic procedures [3, 4], neuro-fuzzy techniques [5, 6], clustering methods [7, 8], genetic algorithms [9-11], fuzzy decision trees [12, 13], and data mining techniques [14, 15].

The antecedent part of fuzzy rules may contain single-dimensional fuzzy sets obtained by partitioning each input dimension. Antecedent fuzzy sets may, e.g., have pre-specified linguistic values with fixed membership functions obtained by homogeneously partitioning each axis of the pattern space [16] or may be purposely defined by domain experts. Alternatively, multi-dimensional antecedent fuzzy sets may be generated by applying a clustering algorithm to sample input-output data [17]. Sometimes, these multi-dimensional antecedent fuzzy sets are projected onto each axis of the input space to improve the interpretability of the clusters produced [18]. In all cases, the output class associated with each fuzzy subset (either grid cell, identified by the partitions on the input dimensions, or cluster) of the pattern space is derived from the training samples belonging to that subset.

Of course, the performance of an FRBC depends on the grain size of the fuzzy partition of the pattern space: a too coarse fuzzy partition may cause many misclassifications while a too fine fuzzy partition may miss to generate fuzzy if-then rules due to lack of training samples in the corresponding areas of the input space. A possible solution is to simultaneously use different partitions with different resolutions at the expense of a high number of fuzzy rules, especially in high-dimensional spaces [19]. Other alternatives are possible. In [20], a hierarchical fuzzy partition is generated independently over each dimension in an ascending way by aggregating fuzzy sets. In [21], a hierarchical fuzzy rule-based classification system is proposed for imbalanced datasets. Basically a finer granularity of the fuzzy partitions is applied in the boundary areas between the classes.

Regarding fuzzy parameter optimization, several techniques have been applied to set the fuzzy system parameters based on the training samples. These include the type and shape of fuzzy membership functions, and the number and structure of fuzzy rules. E.g., Wang and Lee [22] apply the Mapping-Constrained Agglomerative clustering method to identify the cluster configuration of a given dataset for the construction of an initial classifier structure. The linear and nonlinear parameters of the

978-3-901882-56-2@IFIP,2013

classifier are then optimized, respectively, by a recursive least squares algorithm and a modified Levenberg-Marquardt algorithm.

In addition, genetic algorithms (GAs) are often used together with other computational intelligence techniques, to produce intelligent hybrid systems. A GA is an optimization process which starts with a randomly generated initial population of chromosomes, representing candidate solutions to the problem at hand, and evolves toward populations having a better fitness. E.g., in [2], interval-valued fuzzy sets with a post-processing genetic tuning step of their parameters are used to model the linguistic labels. Li and Wang [1] propose a hybrid co-evolutionary GA for learning approximate fuzzy rules, by using a q-nearest neighbor replacement method to coevolve a population of rules, and a local search method. A classifier is built by extracting rules with minimal redundancy from the final population. Setnes and Roubos [9] apply fuzzy clustering to produce an initial TSK fuzzy rule set, then they use a real-coded GA to simultaneously optimize the rule antecedents and the consequents. Ishibuchi *et al.* [23] propose the combination of two fuzzy genetic learning approaches into a single hybrid algorithm for designing FRBCs. Abonyi *et al.* [24] use a decision tree-based initialization of the FRBC for feature selection and initial partitioning of the input domains. The initial fuzzy classifier is optimized by a similarity-driven rule reduction and a multi-objective GA.

In this paper we introduce an easy-to-use approach for efficiently extracting fuzzy rules from available data [25, 26] and we propose its application to three real-world datasets related to energy systems applications. The fuzzy system is obtained exploiting a hierarchical scheme, as a combination of fuzzy models built (employing the fuzzy rule-based classifier `frbc` [27]) on input domain regions increasingly smaller, according to a multi-level grid-like partition. Only the necessary partitions are built, in order to avoid the explosion of the number of rules with the increase of the hierarchical level. The training technique available in `frbc` for the generation of the rule base is the Wang and Mendel method extended to classification problems, an adaptation of the well-known namesake method for regression problems [28]. The fuzzy reasoning method (FRM) employed in `frbc` is a general model of fuzzy reasoning for combining information provided by different rules and determining the output class. It is an extension, presented in [29], of the fuzzy classifier defined by Kuncheva [30].

Finally, a GA is used to optimize some free parameters of the hierarchically built model, by obtaining a genetic-fuzzy system, called from now on "HFRBC-GA". The validity of the proposed approach has been confirmed by applying it on 10 well-known benchmark datasets. The achieved results compare favorably with those obtained by other authors using different techniques.

To illustrate the proposed approach, we employ three case studies related to energy systems applications. More in detail, we refer to the assessment of the energy produced by a solar photovoltaic (PV) installation and to the evaluation of two different kinds of building's energy consumptions. In fact, recently, thanks also to the growing evolution of technologies, the energy sector has drawn the attention of the research community in proposing useful tools to deal with issues of energy efficiency in buildings and with solar energy production management. In addition, the European Commission has adopted a plan to reduce energy consumption of 20% by 2020 [31, 32], by promoting energy efficiency, and the use of renewable energy sources.

The first case study refers to the estimation of PV energy. A PV installation consists of a series of solar panels which using sunlight energy generate directly usable electricity due to the PV effect. A PV panel (see Fig. 1) is composed in its turn of individual PV cells. Usually several panels are connected together to form a system called PV array, to which an inverter is connected that measures the production power of that array and converts the DC power into AC power, as requested by the electrical network. PV installations are typically used as energy sources for the electric grid. In fact, PV energy is considered a free, clean and inexhaustible energy source. Major issues in electric grid management are efficiency and reliability, which require, among other things, fast and easy understanding by the grid operator of both the electricity demand and the electricity supply (energy production). The manager of a PV plant can gain enough information from the system so as to perform appropriate functional operations for the installation, even if the exact energy production value is not known [33].

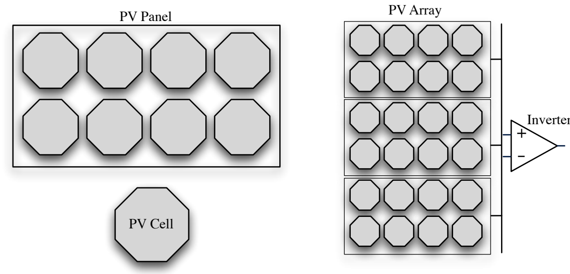

Figure 1. Photovoltaic elements: PV cell, PV panel, and PV array.

The second and third case studies refer to the estimation of energy consumption in buildings due to electric lighting and hot water boiler use. Energy consumption in buildings is one of the fastest growing sectors. It is estimated that the amount of the energy currently consumed in the European buildings is about 40–45% of the total European energy consumption. The potential benefits of estimating energy consumption can be useful for several purposes, ranging from cost reduction, improved energy control, and smarter load scheduling in the electric grid. The paper has the following structure. Section II and III, present, respectively, the proposed hierarchical approach to fuzzy classifiers construction, and the application of the proposed approach to three real-world case studies regarding the energy field. Section IV is devoted to validate our classifier, by comparing it with other fuzzy classifiers present in the literature on 10 well-known benchmark classification problems. Lastly, concluding remarks are provided in Section V.

II. The Proposed Approach

In this section we introduce the proposed hierarchical approach to fuzzy classifier construction.

The methodology consists of a first step, a second iterative step and a final third step. Let us make some general considerations before describing each step in greater detail. Both in the first step and at each iteration of the second step we build a grid, respectively, on the whole input space and on a portion of the input space. Whatever the case, our aim is to find *univocal mapping* areas, i.e., input areas mostly containing patterns associated with the same class label. For each such area, an appropriate number of training samples are randomly extracted and used to generate fuzzy rules that model that area. Since we are interested in collecting training samples according to the real distribution of the available input patterns in relation with each output class, whenever we need to construct a grid in the input portion under consideration we should adopt an ad hoc non-uniform partition, e.g., based on the distribution of the input samples in the feature space. On the other hand, the `frbc` method expects a uniform partition of the input space. Thus, for a good compromise between efficiency and computational cost, we chose to perform a non-uniform grid partitioning of the original input space only in the first step, while we decided to adopt uniform grid partitioning of the relevant input area in all iterations of the subsequent second step. Of course, appropriate scaling will let the non-uniform grid partition correspond to an equivalent uniform partition used by the `frbc` system.

In practice, our objective is to split the input domain into univocal mapping areas with possibly different grain size, and to build a separate set of fuzzy rules to model each such area. Let us now describe more thoroughly the three steps of the methodology.

The methodology can be applied to different kinds of classification problem. More in detail, any number of output classes greater than or equal to two is allowed. Similarly, any number of input features greater than or equal to one is allowed.

A. First Step: First-level Grid Partitioning

In the first step, applied to the original input space, we carry out the following actions.

- i) We apply the *k*-means clustering algorithm [34] separately to each input dimension, with *k* being the number of clusters appropriately chosen.
- ii) We use the separation thresholds between the clusters for:
  - ii.1) building a non-uniform grid in the whole input space, and
  - ii.2) constructing a non-uniform fuzzy partition on each input dimension consisting of *k* membership functions, so as to model each input with linguistic variables.

Since we use the Wang and Mendel method implemented in `frbc`, which builds a uniform fuzzy partition of each input feature space, two more operations must be performed within action ii), namely:

  - ii.3) for each input feature we build a uniform fuzzy partition, consisting of *k* fuzzy sets, using the Wang and Mendel method implemented in `frbc`;
  - ii.4) for each input feature, we use non-uniform scaling to transform the previous uniform partition into the corresponding non-uniform partition (built at stage ii.2): all the feature values are scaled from their original interval to the new interval, maintaining the proportionality.
- iii) We analyze separately each area of the grid previously built in order to discriminate among *insignificant*, *univocal mapping* and *to-subgrid* areas. More precisely:
  - iii.1) an *insignificant* area is any grid area A containing a total number N_A of input samples below a predefined *first-step relevance threshold* RT_1 (the value of RT_1 depends on the specific problem under consideration) and corresponds, e.g., to incompatible or unusual input conditions;
  - iii.2) a *univocal mapping* area A is any non-insignificant area in which there exists a *dominant majority class*, i.e, the class associated with the majority of the samples falling in that area, such that the number N_A^+ of majority class samples is greater than, or equal to, a given percentage, say *first-step dominance percentage* (DP_1), of the numerousness N_A of samples falling in A;
  - iii.3) each non-univocal and non-insignificant grid area is called *to-subgrid* area: each such area will undergo the second iterative step.
- iv) For each univocal mapping area A, a random extraction of $K = \min(perc \cdot N_A^+, S)$ majority class samples is performed, with *perc*, appropriately chosen, representing a percentage of N_A^+, and S, appropriately chosen, being a problem-dependent upper bound of samples of the same class that can be extracted from the same area. The extracted training samples will be used to generate, through `frbc`, the pertinent fuzzy classification rules that model the considered area.
- v) We build the *first-level* fuzzy model by training `frbc` with all the samples extracted from all the univocal mapping areas previously found.

B. Second (Iterative) Step: Deeper-level Grid Partitioning

The second step is applied to each *to-subgrid* area, which has been found either in the first step or at any iteration of the second step itself. For a given *to-subrid* area A we perform the following actions:

- i) we build a uniform hard partition (consisting of *k* intervals) on each input dimension of A, so as to construct a deeper-level uniform grid of the area itself;
- ii) we identify the *insignificant*, *univocal mapping* and *to-subgrid* areas inside the new grid. Similarly to

what done before, first we eliminate from further consideration any insignificant area of the new grid, by using the second-step relevance threshold RT_2^i, with i, $i \geq 1$, representing the iteration number of the second step; then we identify the univocal mapping areas based on the second-step dominance percentage DP_2^i with i, $i \geq 1$, having the same meaning as before; finally, each *to-subgrid* area of the new grid will undergo the second iterative step, thus giving origin to one more iteration. Of course, second-step relevance thresholds RT_2^i, $i \geq 1$, will typically decrease with the increase of the iteration number i, while second-step dominance percentages DP_2^i, $i \geq 1$, may vary according to the iteration number i;

iii) we identify the minimum (hyper)rectangle (see Fig. 2) containing all the samples falling inside the univocal mapping areas included in A; we construct a uniform fuzzy partition, consisting of k membership functions, on each dimension of the (hyper)rectangle; then we generate a *deeper-level fuzzy model* for the (hyper)rectangle by training `frbc` with an appropriate number $K = \min(perc \cdot N_a^+, S)$ of majority class samples extracted from each *univocal mapping* area a related to the hyper(rectangle).

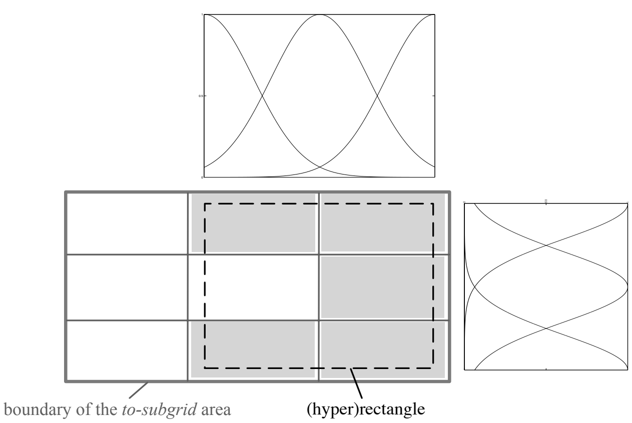

Figure 2. A *to-subgrid* area containing five univocal mapping areas (colored areas), and the minimum (hyper)rectangle (dashed line) considered for building the *deeper-level fuzzy model*.

C. Third Step: Final Fuzzy Model Generation

In the third step, we generate the final fuzzy model, called *merged fuzzy model*, as the union of the *first-level fuzzy model* and all the *deeper-level fuzzy models* built during the hierarchical process, as Fig. 3 exemplifies. Fuzzy models are built on input domain regions increasingly smaller, as the result of the construction of appropriate grids on the pertinent areas of the input domain. The fuzzy sets for each input variable of the *merged fuzzy model* are the union of the fuzzy sets (for that input) of all the models built. We observe that there may be input domain regions modeled by more fuzzy sets (e.g., the larger one built in the first step of analysis and the narrower, and therefore more specific ones, built in the second step of analysis), as shown in Fig. 4. The aim is to exploit the input domain space in an effective way, avoiding unnecessary analysis and thus the generation of too many, irrelevant rules.

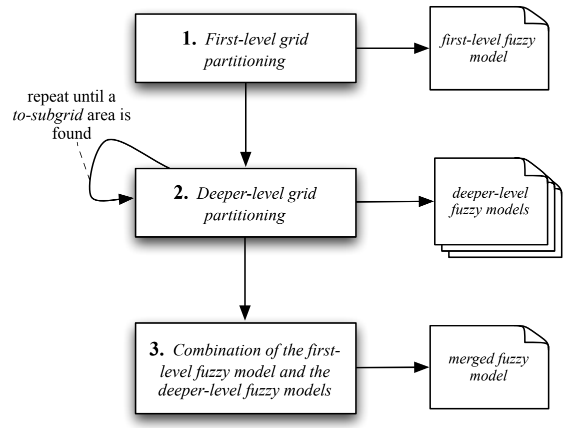

Figure 3. Steps of the proposed methodology and resulting objects.

D. GA-based Parameter Optimization

We apply a GA to optimize the following parameters (i refers to the iteration number of the second step): the relevance thresholds RT_1 and RT_2^i, with $i \geq 1$, the dominance percentages DP_1 and DP_2^i, with $i \geq 1$, the maximum number S of samples extracted from a given grid area (valid for the first step and all iterations of the second step), the minimum rule weight w (valid for the first step and all iterations of the second step), and the rule weight modifiers Δw_1 and Δw_2^i, with $i \geq 1$. In particular, the last two parameters aim, respectively, to control the complexity of the whole rule base, and to enhance/inhibit the influence of the rules of a given step/iteration. The maximum number of iterations is fixed heuristically.

We adopt real-coded chromosomes. The range of possible values of each gene is chosen in heuristic way based on the specific dataset under consideration. When appropriate, integer approximations of real numbers are adopted. The fitness function is the correct classification rate of the fuzzy classifier.

III. Application of the Methodology to Real-world Case Studies

To illustrate the proposed approach we refer to three real-world datasets related to energy systems applications. More in detail, we refer to three experimental case studies seen as three classification problems of i) energy production from a PV installation, ii) electric lighting energy use in office buildings, iii) hot water boiler energy use in office buildings.

To use each of the datasets with the fuzzy classifier we transformed the output numerical values into class labels. For the sake of simplicity, we operated a uniform partition on the output domain by identifying three intervals corresponding to three output classes (*Low*, *Medium*, *High*).

Then, we associated each output value with the energy label corresponding to its interval.

Regarding the free parameters in the methodology for the three case studies examined, in the first step of analysis we adopted, k=3 in the k-means algorithm to obtain the non-uniform partitions of the inputs, and we chose two-sided Gaussian membership functions for the partitions. Indeed they are known to be very accurate, provide complete coverage of the modeled space, and allow easy scaling from the uniform partition to the non-uniform one.

In the second (iterative) step, we used Gaussian membership functions for the fuzzy partitions. Further, we considered the percentage *perc* of majority class samples to extract from each univocal mapping area equal to 70%, both in this first step and in the second step of the methodology.

Regarding the number of hierarchical levels of the methodology, based on heuristic considerations, we considered four possible hierarchical levels, i.e., three iterations of the second step of the methodology. Obviously, the hierarchical levels following the first one are activated only if necessary. At each level, the decomposition of a grid area may actually not generate any fuzzy system, e.g., due to the lack of significant sub-areas found with the decomposition.

The remaining free parameters were set by means of a GA. We used stochastic uniform selection, scattered crossover with probability 0.8, and uniform mutation with probability 0.01. The population consisted of 30 individuals and the maximum number of generations was 300. Thus, a chromosome contains the following real genes: i) the relevance thresholds RT_1 and RT_2^i , i=1, 2, 3, ii) the dominance percentages DP_1 and DP_2^i , i=1, 2, 3, iii) the maximum number of samples extracted S, iv) the minimum rule weight w, and v) the weight modifiers Δw_1 and Δw_2^i, i=1, 2, 3. In addition, for the sake of simplicity, we set $RT_2^3 = RT_2^2 = RT_2^1$ and $DP_2^3 = DP_2^2 = DP_2^1$.

Finally, for each classification problem, we used the following process to calculate the test set accuracy: for each chromosome, we used the values of the genes of that chromosome to perform 30 experiments on 30 different training and test sets randomly generated from the available data, and we computed the fitness as the mean correct classification value on the 30 test sets.

In the following sub-sections we report the results obtained on the examined case studies.

A. PV Energy Production

The data related to the first real-world case study were collected from March to July 2009 and consist of *temperature* of the surface of the solar panel, *solar irradiation*, and *energy production* from the PV installation (output parameter).

We used both the environmental variables, i.e., temperature and irradiation, to estimate the quantity of energy produced by the PV installation.

Fig. 4 shows the first-level and second-level grid partitions on the two-dimensional input domain. As we can see from Fig. 4, areas 3 and 7 are found to be *insignificant* and so they are discarded. Areas 1, 4, 6 and 8 are *univocal mapping* areas. In particular, areas 1 and 4, refer to *low* energy class samples, area 8 refers to *medium* energy class samples, and area 6 refers to *high* energy class samples. Finally, *to-subgrid* areas 2, 5 and 9 are marked for further analysis, thus they are additionally gridded in the first iteration of the second step. The analysis process is repeated iteratively until no area needs to be further divided. For each *to-subgrid* area found at a given analysis level, a partition is built in the following analysis level.

The final fuzzy model is obtained by merging the 23 fuzzy models previously generated in four hierarchical levels of analysis. The final fuzzy model consists of 83 rules and 51 and 38 fuzzy sets, respectively, for the *irradiation* input variable and the *temperature* input variable. Fig. 4 shows (on the top and to the left) a part of the fuzzy sets used to model the linguistic variables.

The best chromosome, resulting from the GA optimization, corresponds to a model which achieved a mean classification performance of 97.38%, with a maximum classification performance of 97.91%. The values of the parameters of the best chromosome are shown in Table I.

B. Electric Lighting Energy Use

The second real-world case study deals with energy consumption in office buildings due to electric lighting energy use. The data, collected from April to June 2010, consist of *solar irradiation*, and *electric lighting energy consumption* (output parameter) expressed in terms of active power. The aim is to derive the class associated with the energy consumption due to electric lighting, knowing basically the external daylight in terms of solar irradiation, and without having to know any information about the building's position and orientation, or the lighting system.

The final fuzzy model is obtained by merging 8 fuzzy models previously generated in four hierarchical levels of analysis. The best model obtained after the GA optimization achieved mean and maximum classification performances of 86.84% and 87%, respectively, and consists of 10 rules and 15 fuzzy sets. The values of the parameters of the best chromosome are shown in Table I.

C. Hot Water Boiler Use

The third real-world case study deals with energy consumption in office buildings due to hot water boiler use, typically for hands washing. The data, collected from April to June 2010, consist of *outdoor temperature*, and *hot water boiler energy consumption* (output parameter) expressed in terms of active power. The aim is to derive the class associated with the energy consumption due to the boiler use, starting from the values of outdoor temperature. Similarly as said before, the system uses only information about external weather conditions and does not need any information about the boiler system.

The final fuzzy model is obtained by merging 7 fuzzy models previously generated in four hierarchical levels of analysis. The best model obtained after the GA optimization achieved mean and maximum classification performances of 88.19% and 88.32%, respectively, and has 4 rules and 15 fuzzy sets. The values of the parameters of the best chromosome are shown in Table I.

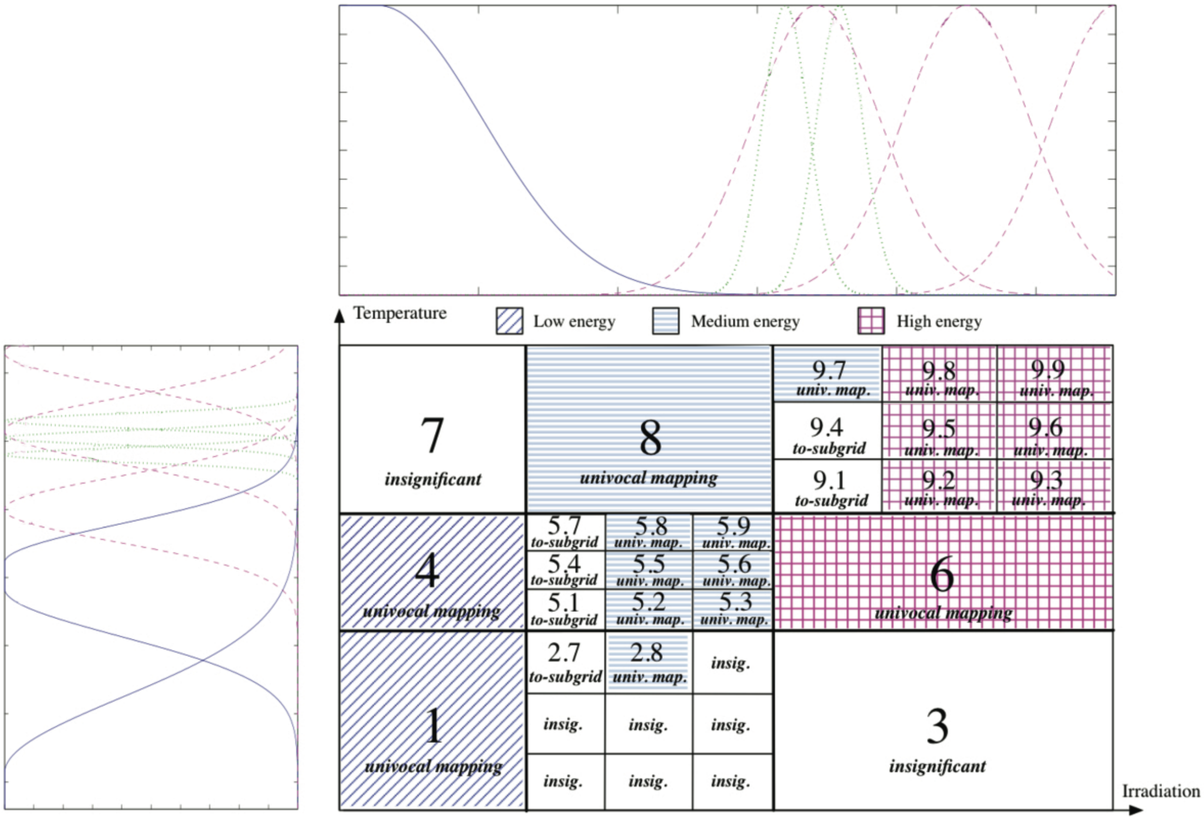

Figure 4. Grid partition of the input space obtained after the application of the first step and the first iteration of the second step of the methodology, and a subgroup of the fuzzy sets employed.

TABLE I. VALUES OF THE GA-OPTIMIZED PARAMETERS

Parameter	**Case study**		
	PV energy	***Hot water boiler***	***Electric lighting***
RT_1	15	4	4
RT_2^i (i = 1, 2, 3)	2	2	3
DP_1	80%	80%	80%
DP_2^i (i = 1, 2, 3)	50%	50%	50%
S	145	26	26
w	0.95	0.91	0.93
Δw_i (i = 1, 2, 3, 4)	{0.45, -0.55, -0.15, 0}	{0.18, -0.06, 0.41, 0}	{0.42, -0.11, -0.05, 0}

IV. EXPERIMENTAL RESULTS ON WELL-KNOWN BENCHMARK DATASETS

This sub-section aims to validate the proposed methodology for building fuzzy classifiers. We applied the fuzzy system built following our approach (HFRBC-GA) to 10 well-known benchmark datasets, whose characteristics are summarized in Table II. All the datasets are available at the UCI machine learning repository [35].

The achieved results are shown in Table III. For each dataset we compared the mean performances of our methodology (HFRBC-GA), achieved with 30 executions of the methodology, with those obtained by other fuzzy classifiers, by choosing the best performing classifier found in the literature. For each model in Table III we report the mean number of rules generated, the number of features actually used (in fact, sometimes we adopted the forward feature selection to decrease the input space dimensionality), the total number of fuzzy sets employed for all features, and the mean test set accuracy. Occasionally, where appropriate, we show also the maximum test set accuracy (in square brackets). Please note that in the table we used the symbol '-' when no information is available, and we introduce a new acronym in quotation marks when the name of the model is not available.

Regarding the datasets Balance, Bupa, Haberman, Ionosphere, Iris, Pima, Sonar, and Wine, our model outperformed the other considered models, most often with fewer features. Regarding the Wisconsin and New Thyroid datasets, our mean results are comparable with those found in the literature (with a difference of 0.08 and 0.52

percentage points, respectively); however, we adopted a lower number of features.

TABLE II. DESCRIPTION OF THE EMPLOYED DATASETS

Dataset	# Samples	# Features	# Classes
Balance	625	4	3
Bupa	345	6	2
Haberman	306	3	2
Ionosphere	351	34	2
Iris	150	4	3
New Thyroid	215	5	3
Pima	768	8	2
Sonar	208	60	2
Wisconsin	683	9	2
Wine	178	13	3

For the sake of completeness, we compared our fuzzy classifier HFRBC-GA with some online classifiers employed by Angelov *et al.* [36], taking into account a selection of datasets from the 10 datasets considered above. More in detail, the datasets considered are: Sonar, Ionosphere, Pima, and Wine. The online classifiers taken into account are the self-evolving FRBC *eClass1*, the incremental C4.5 decision tree-based classifier, and the incremental *k*NN (k-Nearest Neighbor) classifier, with *k*=3. Table IV shows the classification rates of the best online classifier in [36] for each of the 4 datasets, and the difference with the results obtained by HFRBC-GA, which outperforms the online classifiers on all the 4 datasets.

V. CONCLUSIONS

In this paper we have described a hierarchical, genetically optimized, methodology to build a fuzzy classifier by merging fuzzy systems built on input domain regions increasingly smaller, as the result of the creation of appropriate grids on the input domain. The analysis of the input domain space avoids the generation of too many, unnecessary rules and drives to an optimal fuzzy rule base.

To describe the proposed methodology, three experimental case studies related to energy systems applications were carried out. The performance of the proposed approach has also been successfully validated on 10 well-known benchmark classification datasets. The achieved results outperform those obtained by other methods found in the literature.

ACKNOWLEDGMENT

We wish to acknowledge Samares, Via Giuntini, 25, 56023, Navacchio di Cascina, Pisa (Italy), for having provided all the experimental datasets used in this paper.

TABLE III. CLASSIFICATION RESULTS ON SOME BENCHMARK DATASETS AND COMPARISON WITH OTHER FUZZY CLASSIFIERS IN THE LITEARTURE

Dataset	Model	Mean [maximum] test set accuracy (%)	Mean # rules	# Features used	Total # fuzzy sets
Balance	HFRBC-GA	**100**	9	2	6
	FH-GBML [2]	81.12	-	4	-
Bupa	HFRBC-GA	**75.92** [76.72]	83	3	9
	FH-GBML [2]	66.67	-	6	-
Haberman	HFRBC-GA	**87.87** [88.78]	41.7	3	9
	FH-GBML [2]	72.23	-	3	-
Ionosphere	HFRBC-GA	**93.26** [94.67]	17.6	3	9
	“Fuzzy DT” [8]	86.47	3.4	34	-
Iris	HFRBC-GA	**100**	7.1	2	6
	“TSK-GA” [9]	99.4	3	4	12
New Thyroid	HFRBC-GA	95.29 [95.83]	9.6	2	6
	FH-GBML [2]	**95.81**	-	5	-
Pima	HFRBC-GA	**78.31** [80.30]	63.6	3	9
	FH-GBML [2]	75.91	-	8	-
Sonar	HFRBC-GA	**82.67** [85.27]	41.3	2	6
	HGBML [23]	76.30	10	60	-
Wine	HFRBC-GA	**99.41** [100]	10.2	3	9
	SANFIS [22]	99.4	3	13	34
Wisconsin	HFRBC-GA	98.32 [98.44]	14.7	3	9
	HFP [20]	**98.4**	7.8	5	-

TABLE IV. COMPARISON WITH ONLINE CLASSIFIERS

Online classifier	Dataset			
	Sonar	***Ionosphere***	***Pima***	***Wine***
Best model [36]	*eClass1*	Incr. C4.5	*eClass1*	Incr. *k*NN
Classification rate	74.9%	83.88%	74.71%	95.59%
Improvement by HFRBC-GA	+7.77%	+9.38%	+3.6%	+3.82%

REFERENCES

[1] M. Li and Z. Wang, "A hybrid coevolutionary algorithm for designing fuzzy classifiers", Information Sciences, vol. 179, n. 12, 2009, pp. 1970–1983.

[2] J. Sanz, A. Fernández, H. Bustince and F. Herrera, "A genetic tuning to improve the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets: degree of ignorance and lateral position", Int. Journal of Approximate Reasoning, vol. 52, 2011, pp. 751–766.

[3] X.-Z. Wang, Y.-D. Wang, X.-F. Xu, W.-D. Ling and D.-S. Yeung, "A new approach to fuzzy rule generation: fuzzy extension matrix", Fuzzy Sets and Systems, vol. 123, n. 3, 2001, pp. 291–306.

[4] C. Huang and C. Moraga, "Extracting fuzzy if-then rules by using the information matrix technique", Journal of Computer and System Sciences, vol. 70, n. 1, 2005, pp. 26–52.

[5] E.-C.-C. Tsang, X.-Z. Wang and D.-S. Yeung, "Improving learning accuracy of fuzzy decision trees by hybrid neural networks", IEEE Transactions on Fuzzy Systems, vol. 8, n. 5, 2000, pp. 601–614.

[6] S. Mitra and Y. Hayashi, "Neuro-fuzzy rule generation: survey in soft computing framework", IEEE Transactions on Neural Networks, vol. 11, n. 3, 2000 , pp. 748–768.

[7] S. Abe and R. Thawonmas, "A fuzzy classifier with ellipsoidal regions", IEEE Transactions on Fuzzy Systems, vol. 5, n. 3, 1997, pp. 358–368.

[8] J.-A., Roubos, M. Setnes and J. Abonyi, "Learning fuzzy classification rules from labeled data", Information Sciences, vol. 150, 2003, pp. 77–93.

[9] M. Setnes and H. Roubos, "GA-fuzzy modeling and classification: complexity and performance", IEEE Trans. on Fuzzy Systems, vol. 8, n. 5, 2000, pp. 509–522.

[10] E. Zhou and A. Khotanzad, "Fuzzy classifier design using genetic algorithms", Pattern Recognition, vol. 40, 2007, pp. 3401–3414.

[11] A.F. Gómez-Skarmeta, M. Valdés, F. Jiménez and J.G. Marín-Blázquez, "Approximative fuzzy rules approaches for classification with hybrid-GA techniques", Information Sciences, vol. 136, n. 1–4, 2001, pp. 193–214.

[12] Y. Chen, T. Wang, B. Wang and Z. Li, "A Survey of Fuzzy Decision Tree Classifier", Fuzzy Information and Engineering, vol. 1, n. 2, 2009, pp. 149–159.

[13] X.-Z., Wang, E.-C.-C., Tsang and D.-S. Yeung, "A comparative study on heuristic algorithms for generating fuzzy decision trees", IEEE Trans. on Systems, Man and Cybernetics– Part B: Cybernetics, vol. 31, n. 2, 2001, pp. 215–226.

[14] M. De Cock, C. Cornelis and E.E. Kerre, "Elicitation of fuzzy association rules from positive and negative examples", Fuzzy Sets and Systems, vol. 149, n. 1, 2005, pp. 73–85.

[15] M.J. del Jesus, F. Hoffmann, L.J. Navascues and L. Sanchez, "Induction of fuzzy-rule-based classifiers with evolutionary boosting algorithms", IEEE Transactions on Fuzzy Systems, vol. 12, n. 3, 2004, pp. 296–308.

[16] H. Ishibuchi, T. Nakashima and T. Murata, "Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems", IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, vol. 29, n. 5, 1999, pp. 601–618.

[17] S. Abe and M.-S. Lan, "A method for fuzzy rules extraction directly from numerical data and its application to pattern classification", IEEE Transanctions on Fuzzy Systems, vol. 3, n. 1, 1995, pp. 18–28.

[18] H. Roubos and M. Setnes, "Compact and transparent fuzzy models and classifiers through iterative complexity reduction", IEEE Transactions on Fuzzy Systems, vol. 9, n. 4, 2001, pp. 516–524.

[19] H. Ishibuchi, K. Nozaki and H. Tanaka, "Distributed representation of fuzzy rules and its application to pattern classification", Fuzzy Sets and Systems, vol. 52, 1992, pp. 21–32.

[20] S. Guillaume and B. Charnomordic, "Generating an interpretable family of fuzzy partitions from data", IEEE Trans. on Fuzzy Systems, vol. 12, n. 3, 2004, pp. 324–335.

[21] A. Fernández, M.J. del Jesus, and F. Herrera, "Hierarchical fuzzy rule based classification systems with genetic rule selection for imbalanced data-sets", International Journal of Approximate Reasoning, vol. 50, n. 3, 2009, pp. 561–577.

[22] J.-S. Wang and C.S.G Lee, "Self-adaptive neuro-fuzzy inference systems for classification applications, IEEE Trans. on Fuzzy Systems, vol. 10, n. 6, 2002, pp. 790–802.

[23] H. Ishibuchi, T. Yamamoto and T. Nakashima, "Hybridization of fuzzy GBML approaches for pattern classification problems", IEEE Trans. on Systems, Man, and Cybernetics – Part B: Cybernetics, vol. 35, n. 2, 2005, pp. 359–365.

[24] J. Abonyi, J.A. Roubos and F. Szeifert, "Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization", International Journal of Approximate Reasoning, vol. 32, n. 1, 2003, pp. 1–21.

[25] E. D'Andrea and B. Lazzerini, "Computational intelligence techniques for solar photovoltaic system applications", Proc. of Sustainable Internet and ICT for Sustainability (SustainIT), Pisa, Italy, 2012, pp. 1–3.

[26] E. D'Andrea and B. Lazzerini, "A hierarchical approach to multi-class fuzzy classifiers", Expert Systems with Applications, vol. 40, n. 9, 2013, pp. 3828–3840.

[27] M. Cococcioni, E. D'Andrea, and B. Lazzerini, "Providing PRTools with fuzzy rule-based classifiers", Proc. of IEEE International Conference on Fuzzy Systems, Barcelona, Spain, 2010, pp. 1–8.

[28] L.-X. Wang and J.M. Mendel, "Generating fuzzy rules by learning from examples", IEEE Trans. on Systems, Man, and Cybernetics, vol. 22, n. 6, 1992, pp. 1414–1427.

[29] O. Cordón, M.J del Jesus and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems" International Journal of Approximate Reasoning, vol. 20, n. 1, 1999, pp. 21–45.

[30] L.I. Kuncheva, "On the equivalence between fuzzy and statistical classifiers" International Journal of Uncertainty Fuzziness and Knowledge-Based Systems, vol. 4, n. 3, 1996, pp. 245–253.

[31] European Commission, "Energy 2020: A strategy for competitive, sustainable and secure energy", COM(2010), Brussels, 2010.

[32] EREC (European Renewable Energy Council), "Renewable Energy Technology Roadmap: 20% by 2020", Brussels, 2010.

[33] P.A. Gutiérrez, S. Salcedo-Sanz, C. Hervas-Martinez, L. Carro-Calvo, J. Sanchez-Monedero and L. Prieto, "Evaluating nominal and ordinal classifiers for wind speed prediction from synoptic pressure patterns", Proc. of the 11-th International Conf. on Intelligent Systems Design and Applications (ISDA), Córdoba, Spain, 2011, pp. 1265–1270.

[34] J.B. MacQueen, "Some methods for classification and analysis of multivariate pbservations", Proc. of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, 1967, pp. 281–297.

[35] A. Frank and A. Asuncion, UCI Machine Learning Repository (http://archive.ics.uci.edu/ml). Irvine, CA: University of California, School of Information and Computer Science, 2010.

[36] P.P. Angelov and Xiaowei Zhou, "Evolving Fuzzy-Rule-Based Classifiers From Data Streams", IEEE Trans. on Fuzzy Systems, vol.16, n. 6, 2008, pp.1462–1475.