Article

Hybrid Learning for General Type-2 TSK Fuzzy Logic Systems

Mauricio A. Sanchez [1],* [ID]**, Juan R. Castro [1], Violeta Ocegueda-Miramontes [1] and Leticia Cervantes [2]**

[1] School of Engineering, Universidad Autonoma de Baja California, Tijuana 22390, Mexico; jrcastror@uabc.edu.mx (J.R.C.); violeta.ocegueda@uabc.edu.mx (V.O.-M.)
[2] Division of Graduate Studies, Tijuana Institute of Technology, Tijuana 22414, Mexico; lettyy2685@hotmail.com
* Correspondence: mauricio.sanchez@uabc.edu.mx; Tel.: +52-664-682-2790

Abstract: This work is focused on creating fuzzy granular classification models based on general type-2 fuzzy logic systems when consequents are represented by interval type-2 TSK linear functions. Due to the complexity of general type-2 TSK fuzzy logic systems, a hybrid learning approach is proposed, where the principle of justifiable granularity is heuristically used to define an amount of uncertainty in the system, which in turn is used to define the parameters in the interval type-2 TSK linear functions via a dual LSE algorithm. Multiple classification benchmark datasets were tested in order to assess the quality of the formed granular models; its performance is also compared against other common classification algorithms. Shown results conclude that classification performance in general is better than results obtained by other techniques, and in general, all achieved results, when averaged, have a better performance rate than compared techniques, demonstrating the stability of the proposed hybrid learning technique.

Keywords: general type-2 fuzzy logic system; TSK; hybrid learning; principle of justifiable granularity; information granule; classification; granular computing

1. Introduction

Learning techniques in soft computing exist for the purpose of adjusting models so they can accurately represent data in some domain. Although there are various approaches to these learning techniques, we can categorize learning techniques into two groups: hybrid and non-hybrid learning techniques.

A non-hybrid learning technique is composed of a single algorithmic process which achieves the learning of a model, whereas a hybrid learning technique is composed of a sequence of two or more algorithms where in each step a portion of the final model is achieved. Some examples of non-hybrid techniques are a learning algorithm for multiplicative neuron model artificial neural networks [1], an optimized second-order stochastic learning algorithm for neural network training using bounded stochastic diagonal Levenberg–Marquardt [2], design of interval type-2 fuzzy logic systems by utilizing the theory of an extreme learning machine [3], the well-known backpropagation technique for artificial neural networks [4], etc. Yet by combining some of these direct approaches with each other or with other techniques, their performance can greatly improve, such that some steps can compensate performance loss or simply focus on optimizing a portion of the model. Examples of such hybrid models are the scoring criterion for hybrid learning of two-component Bayesian multinets [5], hybrid learning particle swarm optimization with genetic disturbance intended to combat the problem of premature convergence observed in many particle swarm optimization variants [6], a hybrid Monte

Carlo algorithm used to train Bayesian neural networks [7], a learning method for constructing compact fuzzy models [8], etc.

When dealing with raw data where models must be created, transforming such data into more manageable and meaningful information granules can greatly improve how the model performs as well as reducing the computational load of the model. An information granule is a representation of some similar information which can be used to model a portion of some domain knowledge. By forming multiple information granules, these can represent the totality of the information from which data is available; therefore, forming a granular model. Granular computing [9,10] is the paradigm to which these concepts belong.

Information granules which intrinsically support uncertainty can be represented by general type-2 fuzzy sets (GT2 FSs), and in turn these GT2 FSs can be inferred by a general type-2 fuzzy logic system (GT2 FLS) [11]. Although, when dealing with type-2 fuzzy logic systems, they are either in the form of interval type-2 FSs [12] or general type-2 FSs, interval type-2 FSs being a simplification of general type-2 FSs. In simple terms, uncertainty in a GT2 FS is represented by a 3D volume, while uncertainty in an IT2 FS is represented by a 2D area. It is not until recent years that research interest has gained momentum for GT2 FLSs; examples of such published research are fuzzy clustering based on a simulated annealing meta-heuristic algorithm [13], similarity measures α-plane representation [14], hierarchical collapsing method for direct defuzzification [15], a multi-central fuzzy clustering approach for pattern recognitions [16], etc.

Apart from Mamdani FLSs which represent consequents with membership functions, there also exists the representation by linear functions. These FLSs are named Takagi–Sugeno–Kang fuzzy logic systems (TSK FLSs). Examples of TSK FLS usage are evolving crane systems [17], fuzzy dynamic output feedback control [18], analysis of the dengue risk [19], predicting the complex changes of offshore beach topographies under high waves [20], clustering [21], etc.

As published research with GT2 FLS is still very limited, most of it uses Mamdani FLSs, and so far only two published journal papers using a GT2 TSK FLS exist, for controlling a mobile robot [22], and data-driven modeling via a type-2 TSK neural network [23].

In this paper, a proposal of a hybrid learning technique for GT2 TSK FLSs is given, which (1) makes use of the principle of justifiable granularity in order to define a degree of uncertainty in information granules; and (2) a double least square error learning technique is used in order to calculate the parameters for IT2 TSK linear functions. In addition, it is fair to say that at the time of writing of this paper, published research of GT2 TSK FLSs is very limited, therefore this paper contributes to the possibilities that can be achieved by representing consequents with TSK linear functions instead of the more common Mamdani consequents for GT2 FLSs.

This paper is separated into four main sections. First, some background is given which introduces the basic concepts used in the proposed hybrid learning technique; then, the proposed hybrid learning technique is thoroughly described; afterwards, some experimental data is given which defines the general performance of the technique; finally, some concluding remarks are given.

2. Background

2.1. General Type-2 Fuzzy Logic Systems with Interval Type-2 TSK Consequents

A general type-2 fuzzy set (GT2 FS) defined by $\widetilde{\widetilde{A}}$ is represented by $\widetilde{\widetilde{A}} = \left\{ \left((x,u), \mu_{\widetilde{\widetilde{A}}}(x,u) \right) \forall x \in X J_x^u \subseteq [0,1] \right\}$, where X is the Universe of Discourse and $0 \leq \mu_{\widetilde{\widetilde{A}}}(x,u) \leq 1$. In Figure 1, a generic GT2 FS is shown from the primary membership function's perspective. In Figure 2, the same generic GT2 FS is shown but from an isometric view.
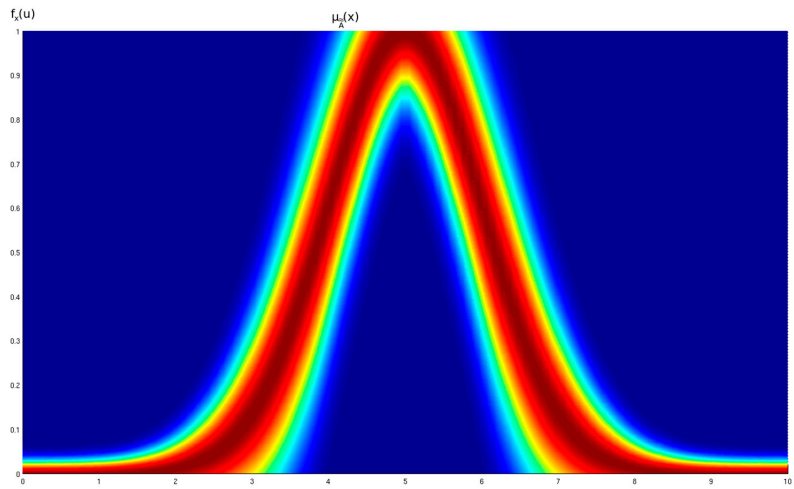
Figure 1. Generic general type-2 fuzzy set (GT2 FS) as shown from the primary function's perspective.
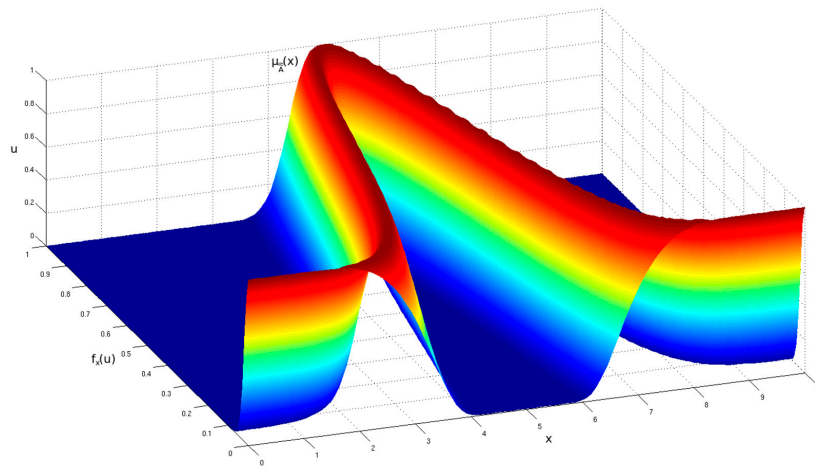

Figure 2. Generic GT2 FS as shown from an isometric view.

The rules of a GT2 FLS are in the form of Equation (1), where R^q is the q-th rule, x_p is the p-th input, $\widetilde{\widetilde{F}}_p^q$ is a membership function on the q-th rule and p-th input, $\widetilde{f}^q$ is an interval type-2 Takagi–Sugeno–Kang (IT2 TSK) linear function on the q-th rule.

$$R^q \text{ IF } x_1 \text{ is } \widetilde{\widetilde{F}}_l^q \text{ and } \dots \text{ and } x_p \text{ is } \widetilde{\widetilde{F}}_p^q, \text{ THEN } \widetilde{f}^q, \text{ where } q = 1, \dots, Q \tag{1}$$

An IT2 TSK linear function [24] $\widetilde{f}^q = \left[f_l^q, f_r^q \right]$ takes the form of Equations (2) and (3), where f_l^q and f_r^q are the left and right switch points of the IT2 TSK linear function on the q-th rule, $c_{q,k}$ is the k-th coefficient on the q-th rule, x_k is the k-th input, $c_{q,0}$ is the constant on the q-th rule, $s_{q,k}$ is the spread k-th coefficient on the q-th rule, and $s_{q,0}$ is the spread on the constant on the q-th rule.

$$f_l^q = \sum_{k=1}^{p} c_{q,k} x_k + c_{q,0} - \sum_{k=1}^{p} |x_k| s_{q,k} - s_{q,0} \tag{2}$$

$$f_r^q = \sum_{k=1}^{p} c_{q,k} x_k + c_{q,0} + \sum_{k=1}^{p} |x_k| s_{q,k} + s_{q,0} \tag{3}$$

2.2. General Type-2 Membership Function Parameterization

The proposed hybrid learning technique depends on the parameterization of a GT2 FS in the form of a Gaussian primary membership function with uncertain mean and Gaussian secondary membership

functions. This GT2 membership function requires four parameters: $\{\sigma, m_1, m_2, \rho\}$, where σ is the standard deviation of the primary membership function, m_1 and m_2 are the left and right centers of the Gaussian membership function with uncertain mean, and ρ is a fraction of uncertainty which affects the support of the secondary membership function. Here, for the sake of simplification, the primary membership function is best represented by the support footprint of uncertainty (FOU) of the primary membership function in the form of an IT2 membership function, as shown in Figure 3. Based on the parameterized structure of the support FOU, the hybrid learning technique performs two type-1 TSK optimizations, as if optimizing two distinct type-1 TSK FLSs.
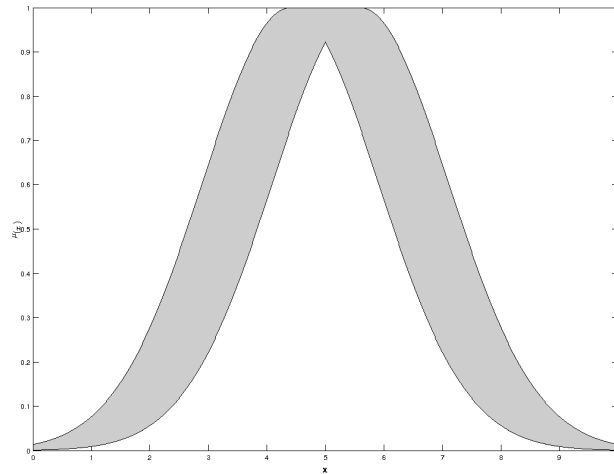

Figure 3. Support of the primary membership function of the used GT2 FS.

The parameterization of the GT2 membership function is as follows. First, the support of the GT2 membership function is created by Equations (4)–(7), using $\{x, u, \sigma, m_1, m_2, \rho\}$, where $x \in X$ on the Universe of Discourse X, and $u \in U$ such that $u \in J_x \subseteq [0, 1]$. Creating an IT2 MF with $\overline{\mu}(x)$ and $\underline{\mu}(x)$, for the upper and lower membership function respectively, as shown in Figure 3.

$$\mu_1(x) = exp\left[-\frac{1}{2}\left(\frac{x - m_1}{\sigma}\right)^2\right] \tag{4}$$

$$\mu_2(x) = exp\left[-\frac{1}{2}\left(\frac{x - m_2}{\sigma}\right)^2\right] \tag{5}$$

$$\overline{\mu}(x) = \begin{cases} \mu_1(x) \ x < m_1 \\ 1 \ m_1 \le x \le m_2 \\ \mu_2(x) \ x > m_2 \end{cases} \tag{6}$$

$$\underline{\mu}(x) = \begin{cases} \mu_2(x) \ x \le \frac{m_1 + m_2}{2} \\ \mu_1(x) \ x > \frac{m_1 + m_2}{2} \end{cases} \tag{7}$$

Afterwards, all parameters required to form the individual secondary membership functions must be calculated, as shown in Equations (8)–(10), where p_x and σ_u are the center and standard deviation of the secondary Gaussian membership function, and ε is a very small number, e.g., 0.000001.

$$p_x = exp\left[-\frac{1}{2}\left(\frac{x - m}{\sigma}\right)^2\right]; \text{ where } m = \frac{m_1 + m_2}{2} \tag{8}$$

$$\delta = \overline{\mu}(x) - \underline{\mu}(x) \tag{9}$$

$$\sigma_u = (1 - \rho)\frac{\delta}{2\sqrt{3}} + \varepsilon \tag{10}$$

Finally, each secondary membership function can be calculated by Equation (11), such that $\widetilde{\mu}(x, u)$ is the secondary function on x. Therefore, forming a complete GT2 membership function would be achieved by calculating for all $x \in X$.

$$\widetilde{\mu}(x, u) = exp\left[-\frac{1}{2}\left(\frac{u - p_x}{\sigma_u}\right)^2\right] \tag{11}$$

### 2.3. Principle of Justifiable Granularity

The purpose of this principle [25] is to specify the optimal size of an information granule where sufficient coverage for experimental data exists while simultaneously limiting the coverage size in order to not overgeneralize. These differences are shown in Figure 4.
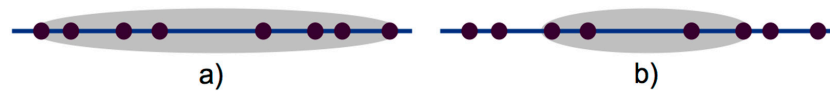
Figure 4. Visual representation of both contradicting objectives in data coverage, where (**a**) complete experimental data coverage is obtained; and (**b**) a limited coverage of experimental data is obtained.

A dual optimization must exist which can consider both objectives, where (1) the information granule must be as specific as possible; and (2) the information granule must have sufficient numerical evidence.

As the length of an information granule is perceived by two delimiting sides of an interval, the dual optimization is performed once per each side. As shown in Figure 5, the left side interval from the Median of the data sample a and the right side interval from the Median of the data sample b creates two intervals to be optimized, where $Med(D)$ is the Median of available data D which initially constructed said information granule.
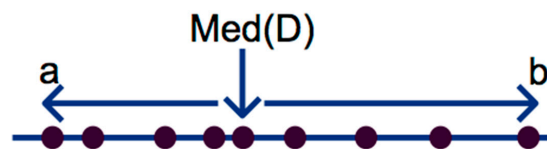
Figure 5. Intervals a and b are optimized from available experimental data for the formation of said information granule, where both lengths start at the median of the information granule's experimental data.

Shown in Equations (12) and (13) are the search equations $V()$ for optimizing a and b respectively, where $V()$ is an integration of the probability density function from $Med(D)$ to all prototypes of a, or b, multiplied by the user criterion for specificity α, where α is a variable which affects the final length of a or b, such that α_0 has the highest experimental data, and α_{max} represents the most specific possible length and has minimal experimental data.

$$V(a) = e^{(-\alpha|Med(D) - a|)}\int_a^{Med(D)} p(x)dx \tag{12}$$

$$V(b) = e^{(-\alpha|b - Med(D)|)}\int_{Med(D)}^b p(x)dx \tag{13}$$

3. Description of the Hybrid Learning Technique for GT2 TSK FLS

The proposed approach, being hybrid in nature, is composed of a sequence of multiple steps, each using different algorithms in order to achieve the final model which is shown in this section.

The hybrid learning technique requires a set of meaningful centers x for the antecedents of the rule base, these can be obtained via any method, such as clustering algorithms; for this paper, a fuzzy c-means (FCM) clustering algorithm [26] was used. Via these cluster centers, subsets belonging to each cluster center are selected through Euclidean distance, where the nearest data point to each cluster center is a member of its subset.

After all subsets are found, cluster coverages can be calculated, i.e., the standard deviation σ, obtained through Equation (14), where $\sigma_{q,k}$ is the standard deviation of the q-th rule and k-th input, x_i is each datum from the subsets previously obtained, $x^c_{q,k}$ is the cluster center for the q-th rule and k-th input, and n is the cardinality of the subset.

$$\sigma_{q,k} = \frac{\sum_{i=1}^{n}\left(x_i - x^c_{q,k}\right)^2}{n-1} \tag{14}$$

Up to now, a Type-1 Gaussian membership function can be formed with $\{\sigma, x\}$. However, the sought end product is a GT2 Gaussian primary membership function with uncertain mean and Gaussian secondary membership functions, which requires the parameters: $\{\sigma, m_1, m_2, \rho\}$. So far, we have calculated σ and partially m_1 and m_2, which are based on x. The following process obtains the remaining required parameters $\{m_1, m_2, \rho\}$ to form GT2 FSs of the antecedents.

To obtain m_1 and m_2, the principle of justifiable granularity is used as a means to heuristically measure uncertainty via the difference of the intervals a and b. This is carried out by extending each information granule to its highest coverage by using the user criterion value of α_0 for each side of the information granule's interval, as described by Equations (12) and (13). When both intervals a and b are obtained, their difference will define the amount of uncertainty which will be used to calculate parameters $\{m_1, m_2\}$, as shown in Equation (15), where τ is a measure of uncertainty for the q-th rule and k-th input.

$$\tau_{q,k} = \left| a_{q,k} - b_{q,k} \right| \tag{15}$$

The obtained value of $\tau_{q,k}$ is used by Equations (16) and (17) to offset the centers $x^c_{q,k}$ of the Gaussian primary membership function by adding uncertainty in the mean, thus obtaining $\{m_1, m_2\}$.

$$m_1^{q,k} = x^c_{q,k} - \tau_{q,k} \tag{16}$$

$$m_2^{q,k} = x^c_{q,k} + \tau_{q,k} \tag{17}$$

For practical reasons, the final missing value of $\{\rho\}$ is set to zero $\rho = 0$ for all membership functions, as it was found that it has no effect on classification performance if other values are set; some experimentation to demonstrate this is shown in Section 4. This ends the parameter acquisition phase for all antecedents in the GT2 TSK FLS.

All IT2 linear TSK consequents are calculated in a two-step process. First, a Least Square Estimator (LSE) algorithm [27,28] is used twice; as the Gaussian primary membership function with uncertain mean is parameterized by a left and right T1 Gaussian membership function on the support FOU, the LSE is applied as if two T1 TSK FLSs existed, using the following sets of parameters: for the left $\left\{\sigma_{q,k}, m_1^{q,k}\right\}$ and right side $\left\{\sigma_{q,k}, m_2^{q,k}\right\}$. When all TSK coefficients φ are obtained, the average of both sets of parameters is used, as shown in Equation (18), where φ_l and φ_r are the coefficient sets for the left and right side respectively. This set C represents all $c_{q,k}$ coefficients of all IT2 TSK linear equations.

$$C = \frac{\varphi_l + \varphi_r}{2} \tag{18}$$

The second and final part of the process for calculating the final spreads $s_{q,k}$ of each coefficient, in set S, which is carried out by measuring the absolute difference between each coefficient set, φ_l and φ_r, as shown in Equation (19).

$$S = |\varphi_l - \varphi_r| \tag{19}$$

A schematic of the proposed hybrid algorithm is shown in Figure 6, where all steps described in this section concentrate the sequence to obtain the antecedents and consequents of the GT2 TSK FLS model, as well as associating certain key steps to their corresponding equations.
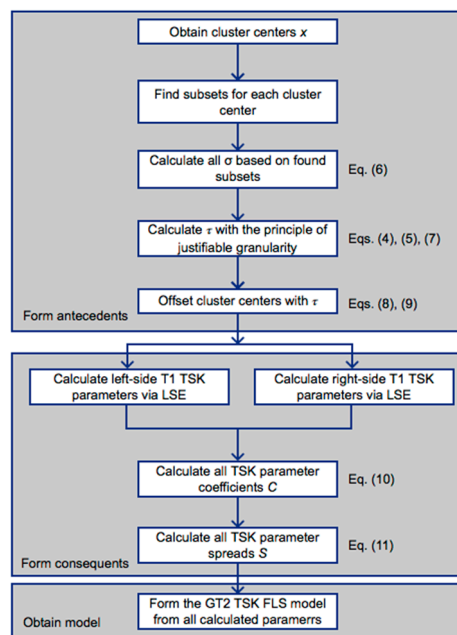
Figure 6. General schematic of the sequence taken by the proposed hybrid algorithm, such that antecedents are calculated first, and consequents afterwards.

4. Experimentation and Results Discussion

A set of various experiments was conducted with classification benchmark datasets in order to explore the effectiveness of the proposed hybrid algorithm. Table 1 shows a compact description of used classification benchmarks [29].

Table 1. Description of used classification benchmarks.

Dataset Name	# of Attributes	# of Classes	# of Samples
Iris	4	3	150
Wine	13	3	178
Breast cancer	9	2	699
Glass identification	9	2	214
Crab gender	6	2	200
Haberman survival	3	2	306
Qualitative bankruptcy	6	2	250
Fertility	9	2	100
Vertebral column	6	3	310
Indian liver patient	10	2	583
Seeds	7	3	210

Experimentation was done using Hold-Out data separation, with 60% randomly selected training data and 40% test data, showing the mean value and standard deviation of 30 execution runs. Concerning the number of rules used per each class, in principle, better model generalization is

usually achieved by reducing the number of rules per class, instead of increasing the number of rules and possibly falling into a case of overfitting [30–32]. For that reason and for simplification purposes, one-rule-per-class was used for all experiments, i.e., results for the iris dataset, which has three classes, were represented by three fuzzy rules, and so on.

Results are shown in Table 2, where values in bold achieved the best performance. Results were compared to Fuzzy C-Means (FCM) [26], Subtractive algorithm [33], Decision Trees, Support Vector Machine (SVM) [34], K-Nearest Neighbors (KNN) [35], and Naïve Bayes [36]; and it must be noted that since the common SVM is designed only for binary classification, it cannot work with datasets which have three or more classes, marked with (-). Performance is measured through total classification percentage, where very good and stable results, in general, are achieved by the proposed hybrid learning technique. Values inside ( ), next to each classification percentage, are the standard deviations for the 30 executions runs which achieved each result, where lower values are better and higher are worse; it can be seen that the proposed hybrid algorithm has a general low variance in the obtained results by means of the calculated standard deviation, yet in the wine dataset it had much more variance when compared to the rest of the techniques.

Although the proposed hybrid technique does not always achieve the best results, it does achieve a better overall performance, as shown in Table 3, where the average across the overall dataset results is shown, such that a higher value means better performance in general, demonstrating that the proposed technique is more stable in general.

Table 2. Results for classification benchmarks.

Dataset	Classification % (std.)						
	Proposed Hybrid Learning Technique	FCM	Subtractive	Decision Tree	SVM	KNN	Naïve Bayes
Iris	95.7880 (1.4738)	85.5004 (4.4088)	96.1394 (1.7889)	95.0559 (3.1016)	-	94.8267 (1.8996)	94.8515 (2.4019)
Wine	87.5556 (8.0889)	67.2222 (6.2849)	74.2222 (13.09)	88.2963 (5.5401)	-	70.2593 (4.4034)	96.5926 (2.1084)
Breast cancer	95.5861 (1.1888)	93.2998 (1.6142)	94.1186 (1.121)	91.6426 (2.4014)	94.9867 (1.2044)	93.4189 (1.5414)	94.7079 (1.3952)
Glass chemical	90.6983 (1.9856)	89.5373 (2.9519)	90.6164 (2.9945)	90.7540 (2.3476)	91.3606 (3.165)	93.0778 (2.8169)	89.8460 (3.055)
Crab gender	95.5000 (1.6132)	50.3000 (3.2641)	90.7667 (3.3211)	79.1333 (4.3366)	94.1000 (2.1804)	83.7000 (3.8411)	63.4667 (10.7027)
Haberman survival	74.4116 (2.1708)	72.4723 (2.5779)	71.0549 (2.8901)	66.4893 (4.6845)	72.2641 (2.7463)	66.7862 (3.8423)	73.8685 (2.8525)
Qualitative bankruptcy	99.2404 (0.6214)	82.3181 (4.1822)	98.1105 (1.4148)	98.4230 (1.0181)	98.6156 (0.639)	99.1158 (0.9828)	98.0451 (1.04)
Fertility	87.1333 (2.6261)	86.9333 (3.2582)	80.2667 (4.3676)	82.4667 (7.2603)	86.6000 (3.8721)	80.2000 (5.3653)	85.0000 (3.4157)
Vertebral column	80.8676 (2.9845)	61.7257 (3.8218)	80.8731 (2.8185)	80.2976 (3.3082)	-	80.0110 (3.1363)	77.0305 (3.6083)
Indian liver patient	70.5602 (3.6878)	70.4214 (2.1655)	68.7966 (2.7519)	66.2205 (2.8477)	70.5329 (2.1469)	64.6039 (2.4738)	57.2955 (3.5323)
Seeds	94.3082 (2.1189)	89.3711 (3.1082)	93.1132 (2.9531)	90.5975 (4.0337)	-	89.7484 (2.5469)	90.8805 (2.8962)

Table 3. Overall performance of tested techniques with datasets.

| Proposed Hybrid Learning Technique	FCM	Subtractive	Decision Tree	SVM	KNN	Naïve Bayes
88.3317	77.1910	85.2798	84.4887	86.9228	83.2498	83.7804

In Table 4, an experiment to demonstrate that the value of ρ has no effect whatsoever in the classification performance of the proposed hybrid learning technique is shown. Two datasets were chosen at random with 60% training data and 40% testing data, with $\rho = [0, 0.5, 1, 2]$. To achieve a truer experiment when comparing chosen ρ values, the exact same training data and testing data was used, i.e., with each execution run, data was not randomly separated into a 60/40 partition; instead, the partition was fixed with exact data in each experiment, and only the value of ρ was changed.

Table 4. Demonstration that varying ρ values have no effect on classification performance.

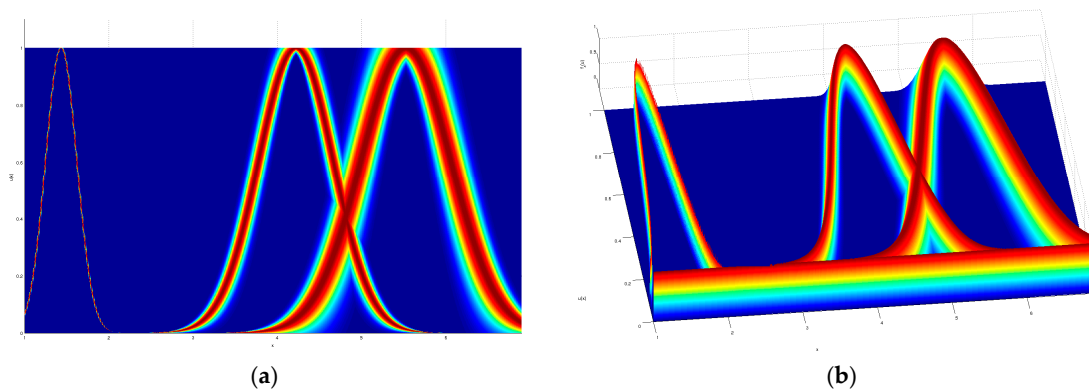| Dataset	ρ			
	0	**0.5**	**1**	**2**
Breast cancer	94.6939	94.6939	94.6939	94.6939
Qualitative bankruptcy	98.7342	98.7342	98.7342	98.7342

As a visual example, the input partition for the iris dataset modeled by a GT2 TSK FLS which obtained the given result in Table 2, is shown in Figures 7–10, where a top and orthogonal view can be seen; in all accounts, the amount of uncertainty within each membership function is quite contrasting, as there are membership functions with barely any uncertainty and also membership functions with quite a lot of uncertainty.
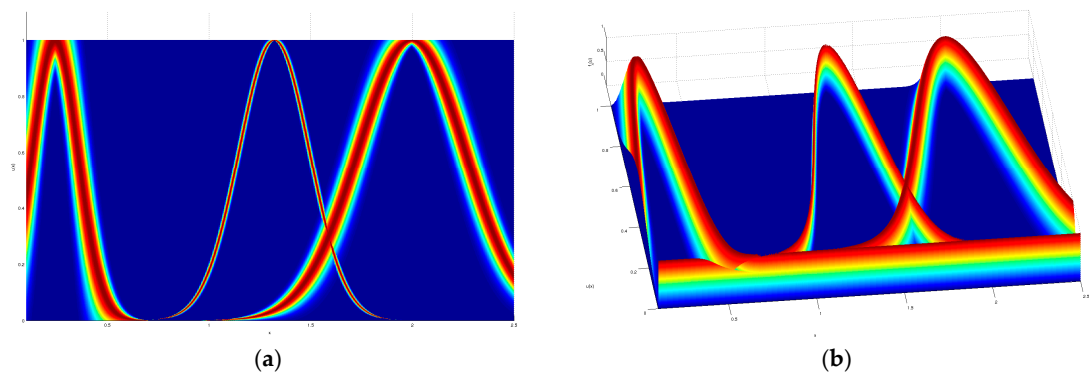

(**a**) (**b**)
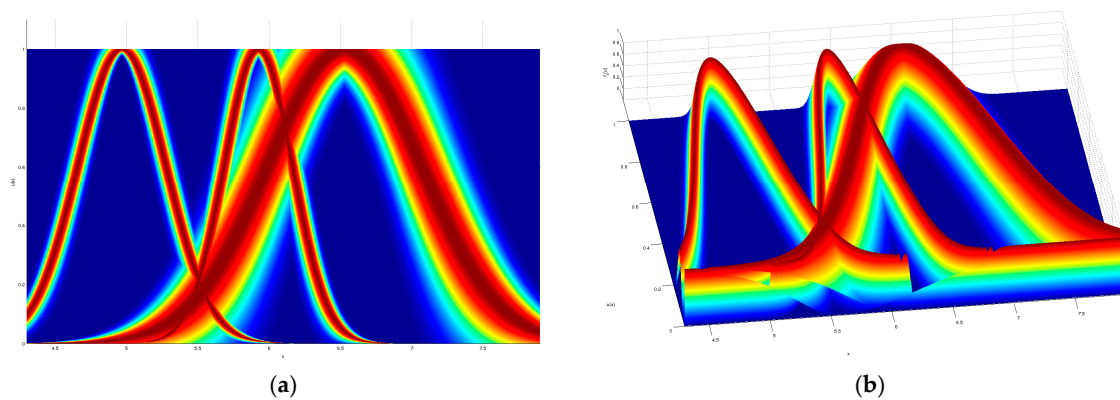
Figure 7. Input partition of the GT2 TSK fuzzy logic system (FLS) for the first input of the iris dataset, where (**a**) shows a top view of the GT2 membership functions; and (**b**) shows an orthogonal view of the same GT2 membership functions.

(**a**) (**b**)
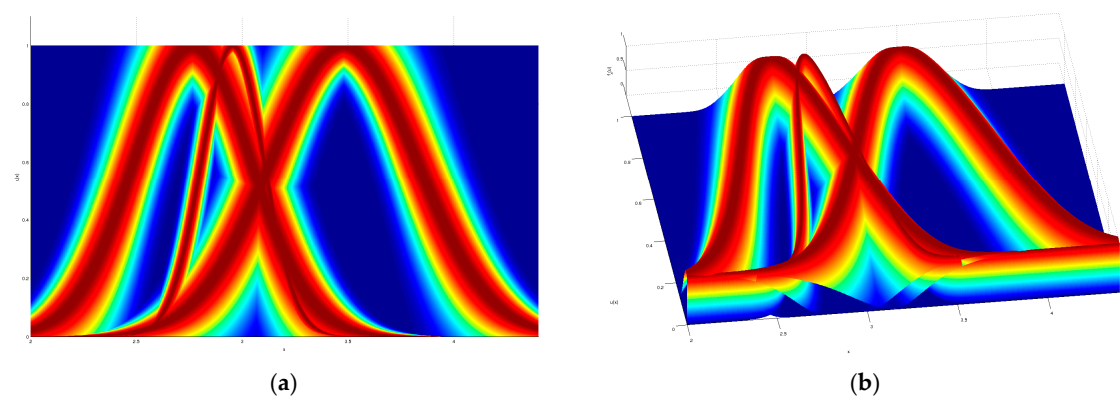
Figure 8. Input partition of the GT2 TSK FLS for the second input of the iris dataset, where (**a**) shows a top view of the GT2 membership functions; and (**b**) shows an orthogonal view of the same GT2 membership functions.

Figure 9. Input partition of the GT2 TSK FLS for the third input of the iris dataset, where (**a**) shows a top view of the GT2 membership functions; and (**b**) shows an orthogonal view of the same GT2 membership functions.

Figure 10. Input partition of the GT2 TSK FLS for the fourth input of the iris dataset, where (**a**) shows a top view of the GT2 membership functions; and (**b**) shows an orthogonal view of the same GT2 membership functions.

It must also be noted that there are a couple of variables where classification performance could be improved. First, as the initial hybrid learning technique requires prototype centers to begin constructing the GT2 FSs around them, if better prototypes are found, then the classification performance is also bound to improve; for the included experimentation in this paper, a FCM clustering algorithm was used to obtain the initial prototypes, yet other techniques could be used to improve the final classification performance by providing better quality initial prototypes. As is known, different techniques perform differently with each dataset, therefore by changing this part of the proposed hybrid algorithm, better results could be expected. Second, in this paper, GT2 FSs in the form of Gaussian primary membership functions with uncertain mean and Gaussian secondary membership functions were used, meaning that other GT2 FSs could be used; this, in itself, is worth exploring in future research, as different FSs could greatly improve the representation of information granules and therefore improve the quality of the fuzzy model. Third, a dual LSE application technique was used to calculate IT2 TSK linear function parameters for all consequents, where other more robust algorithms could be used to further improve the general performance of the model, e.g., Recursive Least Squares (RLS) algorithm.

One of the qualities of the proposed hybrid approach, as shown through experimentation, is the stability inherent in FLSs in general, especially in GT2 FLSs, where the integrated handling of uncertainty in its model permits less variance in achieved performance when compared to other classifying techniques. By using information granules which support varying degrees of uncertainty,

as acquired by the same data which formed it, changing patterns in data have less of an effect on the performance of the fuzzy model created by the proposed hybrid technique.

5. Conclusions

The work proposed in this paper is an initial exploration into the effectiveness of GT2 TSK FLSs for use in classification scenarios. Due to the complexity of GT2 FLSs in general, a hybrid learning technique is introduced. As a result of using a hybrid learning algorithm, a sequence of various stages takes place in order to obtain the final fuzzy granular model; in the first stage, initial prototypes must be acquired from a sample of data; this can be obtained through various means, such as clustering algorithms, providing the flexibility of using any technique which may acquire these initial prototypes with improved quality; in the second stage, some level of uncertainty is defined through the principle of justifiable granularity, where differences between both intervals, *a* and *b*, for each information granule, depict how the spread of data measures uncertainty. The highest coverage is used in both intervals to simplify the information granule's coverage, yet it could be possible to achieve better performance by identifying an optimal value between $[\alpha_0, \alpha_{max}]$ rather than just using α_0 equally for all information granules. Finally, the calculation stage of the IT2 TSK linear function parameters is a direct method reliant on the previous stage which obtains results via a dual application of the LSE algorithm, after which these two sets of parameters are joined to the final required parameters to finish forming the GT2 TSK FLS, where other more precise learning techniques should yield much better parameters for improved model quality.

Experimentation gave a quick view of the general quality of these GT2 TSK FLS models, where a degree of stability was achieved in contrast to other more common classification algorithms. Research into GT2 TSK FLS is still scarce, and this paper showed some of the benefits of model quality, performance, and stability, that this type of system can achieve.

Author Contributions: Mauricio A. Sanchez conceived and designed this research as well writing this paper, Juan R. Castro also conceived and designed this research, Violeta Ocegueda-Miramontes and Leticia Cervantes performed all experiments and performed the discussion of results.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bas, E.; Uslu, V.R.; Egrioglu, E. Robust learning algorithm for multiplicative neuron model artificial neural networks. *Expert Syst. Appl.* **2016**, *56*, 80–88. [CrossRef]
2. Liew, S.S.; Khalil-Hani, M.; Bakhteri, R. An optimized second order stochastic learning algorithm for neural network training. *Neurocomputing* **2016**, *186*, 74–89. [CrossRef]
3. Hassan, S.; Khosravi, A.; Jaafar, J.; Khanesar, M.A. A systematic design of interval type-2 fuzzy logic system using extreme learning machine for electricity load demand forecasting. *Int. J. Electr. Power Energy Syst.* **2016**, *82*, 1–10. [CrossRef]
4. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
5. Carvalho, A.M.; Adão, P.; Mateus, P. Hybrid learning of Bayesian multinets for binary classification. *Pattern Recognit.* **2014**, *47*, 3438–3450. [CrossRef]
6. Liu, Y.; Niu, B.; Luo, Y. Hybrid learning particle swarm optimizer with genetic disturbance. *Neurocomputing* **2015**, *151*, 1237–1247. [CrossRef]
7. Kocadağlı, O. A novel hybrid learning algorithm for full Bayesian approach of artificial neural networks. *Appl. Soft Comput.* **2015**, *35*, 52–65. [CrossRef]
8. Zhao, W.Q.; Niu, Q.; Li, K.; Irwin, G.W. A hybrid learning method for constructing compact rule-based fuzzy models. *IEEE Trans. Cybern.* **2013**, *43*, 1807–1821. [CrossRef] [PubMed]
9. Zadeh, L.A. Fuzzy sets and information granularity. In *Advances in Fuzzy Set Theory and Applications*; North-Holland Publishing Company: Amsterdam, The Netherlands, 1996; pp. 3–18.

10. Bargiela, A.; Pedrycz, W. Toward a theory of granular computing for human-centered information processing. *IEEE Trans. Fuzzy Syst.* **2008**, *16*, 320–330. [CrossRef]

11. Mendel, J. General type-2 fuzzy logic systems made simple: A tutorial. *IEEE Trans. Fuzzy Syst.* **2013**, *22*, 1162–1182. [CrossRef]

12. Mendel, J.M.; John, R.I.; Liu, F. Interval type-2 fuzzy logic systems made simple. *IEEE Trans. Fuzzy Syst.* **2006**, *14*, 808–821. [CrossRef]

13. Doostparast Torshizi, A.; Fazel Zarandi, M.H. Alpha-plane based automatic general type-2 fuzzy clustering based on simulated annealing meta-heuristic algorithm for analyzing gene expression data. *Comput. Biol. Med.* **2015**, *64*, 347–359. [CrossRef] [PubMed]

14. Hao, M.; Mendel, J.M. Similarity measures for general type-2 fuzzy sets based on the -plane representation. *Inf. Sci.* **2014**, *277*, 197–215. [CrossRef]

15. Doostparast Torshizi, A.; Fazel Zarandi, M.H. Hierarchical collapsing method for direct defuzzification of general type-2 fuzzy sets. *Inf. Sci.* **2014**, *277*, 842–861. [CrossRef]

16. Golsefid, S.M.M.; Zarandi, M.H.F.; Turksen, I.B. Multi-central general type-2 fuzzy clustering approach for pattern recognitions. *Inf. Sci.* **2016**, *328*, 172–188. [CrossRef]

17. Precup, R.-E.; Filip, H.-I.; Rădac, M.-B.; Petriu, E.M.; Preitl, S.; Dragoş, C.-A. Online identification of evolving Takagi–Sugeno–Kang fuzzy models for crane systems. *Appl. Soft Comput.* **2014**, *24*, 1155–1163. [CrossRef]

18. Klug, M.; Castelan, E.B.; Leite, V.J.S.; Silva, L.F.P. Fuzzy dynamic output feedback control through nonlinear Takagi–Sugeno models. *Fuzzy Sets Syst.* **2015**, *263*, 92–111. [CrossRef]

19. Silveira, G.P.; de Barros, L.C. Analysis of the dengue risk by means of a Takagi–Sugeno-style model. *Fuzzy Sets Syst.* **2015**, *277*, 122–137. [CrossRef]

20. Kim, Y.; Kim, K.H.; Shin, B.-S. Fuzzy model forecasting of offshore bar-shape profiles under high waves. *Expert Syst. Appl.* **2014**, *41*, 5771–5779. [CrossRef]

21. Sanchez, M.A.; Castillo, O.; Castro, J.R.; Melin, P. Fuzzy granular gravitational clustering algorithm for multivariate data. *Inf. Sci.* **2014**, *279*, 498–511. [CrossRef]

22. Sanchez, M.A.; Castillo, O.; Castro, J.R. Generalized Type-2 Fuzzy Systems for controlling a mobile robot and a performance comparison with Interval Type-2 and Type-1 Fuzzy Systems. *Expert Syst. Appl.* **2015**, *42*, 5904–5914. [CrossRef]

23. Yeh, C.Y.; Jeng, W.H.R.; Lee, S.J. Data-based system modeling using a type-2 fuzzy neural network with a hybrid learning algorithm. *IEEE Trans. Neural Netw.* **2011**, *22*, 2296–2309. [CrossRef] [PubMed]

24. Mendel, J. Unnormalized interval type-2 TSK FLSs. In *Uncertain Rule-Based Fuzzy Logic System: Introduction and New Directions*; Prentice-Hall: Upper Saddle River, NJ, USA, 2001; p. 555.

25. Pedrycz, W. The principle of justifiable granularity and an optimization of information granularity allocation as fundamentals of granular computing. *J. Inf. Process. Syst.* **2011**, *7*, 397–412. [CrossRef]

26. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Springer: Boston, MA, USA, 1981.

27. Jang, J.S.R. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* **1993**, *23*, 665–685. [CrossRef]

28. Jang, J.-S.R. Fuzzy modeling using generalized neural networks and Kalman filter algorithm. In Proceedings of the Ninth National Conference on Artificial Intelligence, Anaheim, CA, USA, 14–19 July 1991; AAAI Press: Palo Alto, CA, USA, 1991; pp. 762–767.

29. Frank, A.; Asuncion, A. *UCI Machine Learning Repository*; University of California: Irvine, CA, USA, 2010.

30. Wang, X.Z.; Dong, C.R. Improving generalization of fuzzy IF–THEN Rules by maximizing fuzzy entropy. *IEEE Trans. Fuzzy Syst.* **2009**, *17*, 556–567. [CrossRef]

31. Jin, Y.; Von Seelen, W.; Sendhoff, B. On generating FC^3 fuzzy rule systems from data using evolution strategies. *IEEE Trans. Syst. Man Cybern. Part B* **1999**, *29*, 829–845. [CrossRef]

32. Cawley, G.C.; Talbot, N.L.C. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.

33. Chiu, S.L. Fuzzy model identification based on cluster estimation. *J. Intell. Fuzzy Syst.* **1994**, *2*, 267–278.

34. Cortes, C.; Vapnik, V. Support vector machine. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

35. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **1992**, *46*, 175–185. [CrossRef]

36. John, G.H.; Langley, P. Estimating continuous distributions in Bayesian classifiers. In Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, Montréal, QC, Canada, 18–20 August 1995; Morgan Kaufmann: Burlington, MA, USA, 1995; pp. 338–345.