

# Uncertainty-Based Information Granule Formation

Mauricio A. Sanchez, Oscar Castillo and Juan R. Castro

**Abstract** A new technique for forming information granules is shown in this chapter. Based on the theory of uncertainty-based information, an approach is proposed which forms Interval Type-2 Fuzzy information granules. This approach captures multiple evaluations of uncertainty from taken samples and uses these models to measure the uncertainty from the difference in these. The proposed approach is tested through multiple benchmark datasets: iris, wine, glass, and a 5th order curve identification.

## 1 Introduction

Granular computing is concerned with how information is grouped together and how these groups can be used to make decisions [1, 2]. It is inspired by how human cognition manages information. Granular computing is used to improve the final representation of information models by forming information granules which better adapt to the known information. Although granular computing expresses information models, more commonly known as information granules, it can use a variety of representations to express such granules, which could be rough sets [3], quotient space [4], shadowed sets [5], fuzzy sets [6, 33–35], etc.

---

M. A. Sanchez · J. R. Castro  
Autonomous University of Baja California, Tijuana, Mexico  
e-mail: mauricio.sanchez@uabc.edu.mx

J. R. Castro  
e-mail: jrcastror@uabc.edu.mx

O. Castillo (✉)  
Tijuana Institutes of Technology, Tijuana, Mexico  
e-mail: ocastillo@tectijuana.mx

Information granules are representations of similar information which can be used for a purpose, typically to model a portion of information. Forming fuzzy information granules is not new, knowing that many representations can be used there have been many approaches which try to solve this: via their relationships [7], optimization of time granularity [8], information granulation [9], with RBF neural networks [10], Interval Type-2 Fuzzy granules [11], non-homogeneous General Type-2 Fuzzy granules [12], etc.

This chapter proposes an approach to information granule formation by capturing, through samples, evaluations of uncertainty where their difference is a direct measure of uncertainty which is used to form Interval Type-2 Fuzzy information granules [25–32].

This chapter is organized as follows: Sect. 2 describes the proposed approach as well as its motivation. Section 3 shows benchmark results alongside the discussion. Finally, Sect. 4 concludes the document.

## **2 Uncertainty-Based Information Granule Formation Methodology Description**

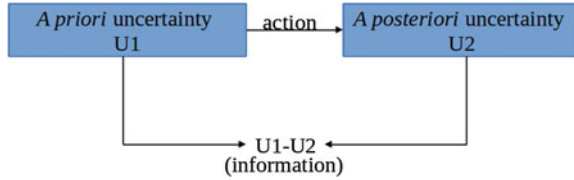
To first understand the main methodology, a review of the motivation is necessary, as it describes the basis for the proposed approach. First, the basis for the proposed approach, which is the theory of uncertainty-based information [13, 14] is described; then, evaluations of uncertainty [15, 16] are described, which defines functions that represent uncertainty measures.

### ***2.1 Uncertainty-Based Information***

The concept of uncertainty is closely related to the concept of information. The fundamental characteristic of this relation is that involved uncertainty from any problem-solving situation is a result of information deficiency pertaining to the system within which the situation is conceptualized. This information could be incomplete, imprecise, fragmentary, unreliable, vague, or contradictory.

With the assumption that a certain amount of uncertainty can be measured from a problem-solving situation it is possible that a mathematical theory can be formed.

With another assumption that this amount of uncertainty is reduced by obtaining relevant information as a result of some action (e.g. obtaining experimental results, observing new data, etc.), the amount of obtained information by the action can be measured by that amount of reduced uncertainty. That is, the amount of information related to a given problem-solving situation that is obtained through some action is measured by the difference between a priori uncertainty and a posteriori uncertainty.



**Fig. 1** Diagram of the behavior of the uncertainty-based information where uncertainty is reduced by the difference between two uncertain models of the same information

In Fig. 1, the shown diagram represents the general idea of the behavior of uncertainty-based information; where a reduction of uncertainty can be obtain by the difference of two uncertain models of the same information. That is, the a priori uncertainty model is obtained with a first sample of information, where as the posteriori uncertainty model is obtained with a second sample of information related to the same problem-solving situation.

### 2.2 Evaluations of Uncertainty

To capture uncertainty, there are two fundamental types of evaluations: Type A and Type B.

Through repeated measurements, an average measured value can infer a standard deviation which forms a Gaussian distribution function, where this functions is a Type A evaluation of uncertainty.

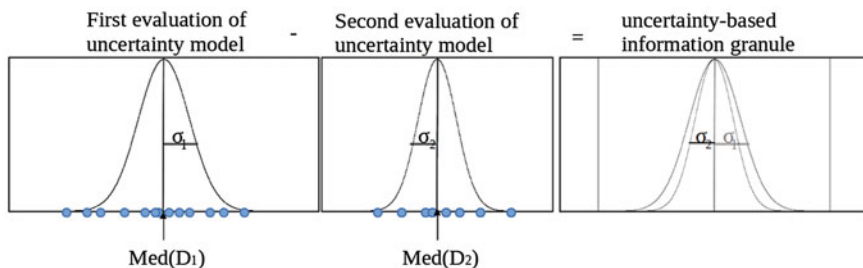
Type B evaluations of uncertainty are represented by a rectangular probability distribution, in other words, a specified interval where the measurements are known to lie in.

### 2.3 Uncertainty-Based Information Granule Formation

Taking inspiration on uncertainty-based information, this can be interpreted in a manner which forms higher-type information granules where uncertainty can be captured and measured and build Interval Type-2 Fuzzy information granules.

A sample of information can build a model with uncertainty from the complete source of information; this is, since it is impossible to know the complete truth of any given situation, uncertainty will always exist in any sample information which may be taken from it.

Through a first sample of information ( $D_1$ ), an uncertain model (evaluation of uncertainty) can be created. Through a second sample of information ( $D_2$ ), another similar uncertain model can be also created. These two models of uncertainty are analogous to the models in the theory of uncertainty-based information, a priori and posteriori uncertainty models.



**Fig. 2** Explanatory diagram of how the proposed approach measures and defines the uncertainty, and forms an IT2 Fuzzy set with such uncertainty

In a direct comparison with the theory of uncertainty-based information, the proposed approach does not reduce the uncertainty in the model, instead it measures and defines it to be able to use it in an information granule and have an improved representation of the information. The proposed approach is shown in Fig. 2, where a first sample of information obtains an evaluation of uncertainty, in the form of a Gaussian function, or Type-1 Gaussian membership function; and a second sample of information obtained another similar evaluation of uncertainty, of the same form. A difference is found between these two Gaussian membership functions defining the Footprint of Uncertainty (FOU), thus obtaining an IT2 Fuzzy information granule. Here there are three possibilities: (1) the first Gaussian membership function has an  $\sigma$  which is larger than the second; (2) the second Gaussian membership function has an  $\sigma$  which is larger than the first; and (3) the  $\sigma$  from both Gaussian membership functions are the same. For 1 and 2, the FOU which is created defines some uncertainty which has been measured and can now be used by the IT2 Fuzzy System; and for 3, since no uncertainty was measured a T1 Fuzzy Set is created.

To show the viability of the proposed approach in that it captures uncertainty and forms IT2 Fuzzy information granules, an algorithm was created that would allow for results to be obtained. The following steps define the algorithms:

1. Obtain rules and centers. These can be obtained through any clustering algorithm, for the experimental case in this chapter the subcluster algorithm [17] was used.
2. Through a first sample of information ( $D_1$ ), all  $\sigma_1$  for all centers are calculated. These were found by calculating the Euclidean n-space distance between each data point and all centers, where the shortest distance defines to which center does that point belong to, afterwards having a set of data points for each cluster, a standard deviation was calculated as to form an evaluation of uncertainty in the form of a Gaussian membership function. For the case of testing, a random sample comprised of 40 % of the dataset was used.
3. Through a second sample of information ( $D_2$ ), in the same manner as the previous step, all  $\sigma_2$  for all centers are calculated. A random sample comprised of another 40 % of the dataset was used for this step.

4. Form the IT2 Fuzzy Gaussian information granules as proposed. This only builds the antecedents of a complete IT2 Fuzzy System.
5. The consequents are finally optimized via an evolutionary algorithm, obtaining a complete IT2 Fuzzy System which can be used to acquire results. For this chapter, Interval Takagi-Sugeno-Kang (TSK) [18, 19] consequents were used, they were optimized via a Cuckoo Search algorithm [20].

The next section uses this algorithm to obtain results.

### 3 Experimental Results and Discussion

For experimental tests, four datasets were used: iris, wine, glass, available from the UCI dataset repository [21], and a 5th order polynomial curve. Where the iris dataset, has 4 input features (petal length, petal width, sepal length, and sepal width), and 3 outputs (iris setosa, iris virginica, and iris versicolor). With 50 samples of each flower type, with a total of 150 elements in the dataset. The wine dataset, with 13 input features of different constituents (Alcohol, malic acid, ash, alkalinity of ash, magnesium, total phenols, flavanoids, nonflavanoid phenols, proanthocyanins, color intensity, hue, OD280/OD315 of diluted wines, and proline) identifying 3 distinct Italian locations where the wine came from. With 59, 71, and 48 elements respectively in each class, for a total of 178 elements in the whole dataset. The glass identification dataset, has 9 input variables (refractive index, sodium, magnesium, aluminum, silicon, potassium, calcium, barium, and iron), and 7 classes (building windows float processed, building windows non float processed, vehicle windows float processed, containers, tableware, and headlamps). With 70, 76, 17, 13, 9, and 29 elements respectively in each class, for a total of 214 elements in the whole dataset.

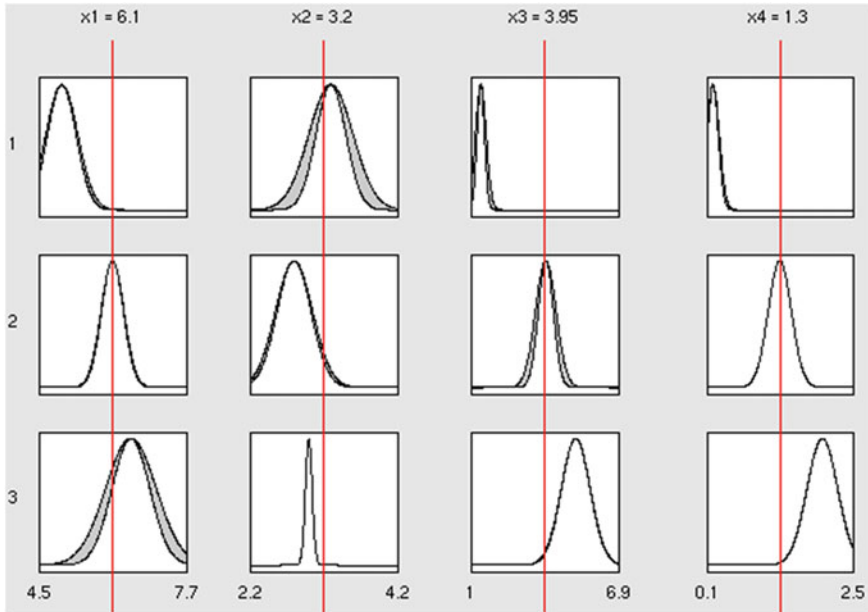
#### 3.1 Experimental Results

On Table 1, the obtained results are shown, where the after 30 execution runs for each dataset were made to obtain a minimum, maximum, mean, and standard deviation for each dataset.

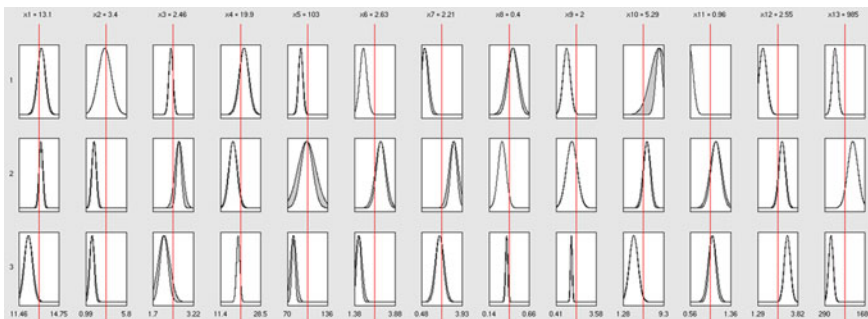
The following Figs. 3, 4, 5, 6 show one sample of the formed IT2 Fuzzy information granules of each dataset: iris, wine, glass, and 5th order polynomial, respectively.

**Table 1** Obtained results for the chosen datasets

	Classification accuracy %			RMSE error
	Iris	Wine	Glass	5th order polynomial
Min	86.66	88.88	44.18	0.36
Mean	93.99	93.05	68.43	0.69
Max	100	97.22	86.04	1.14
Std	5.164	5.982	15.592	0.181



**Fig. 3** Sample of the formed IT2 Fuzzy information granules for the Iris dataset



**Fig. 4** Sample of the formed IT2 Fuzzy information granules for the Wine dataset

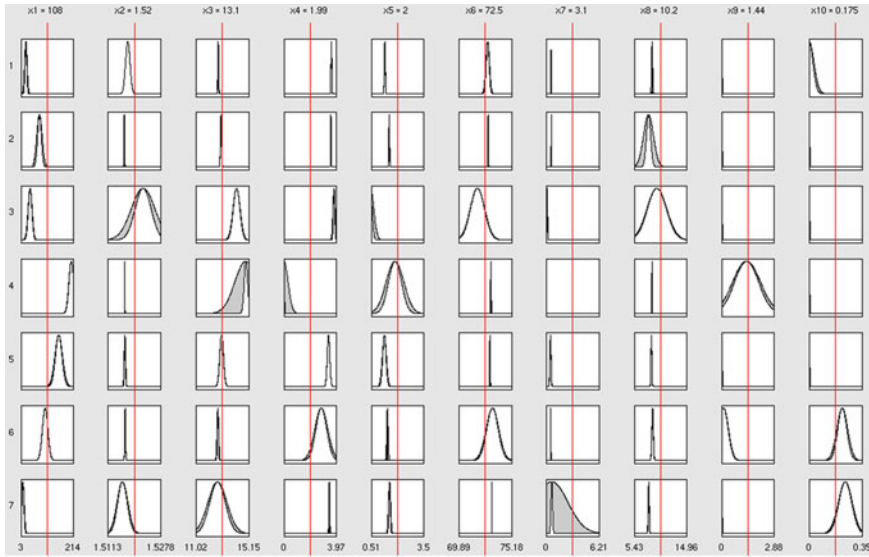


Fig. 5 Sample of the formed IT2 Fuzzy information granules for the Glass dataset

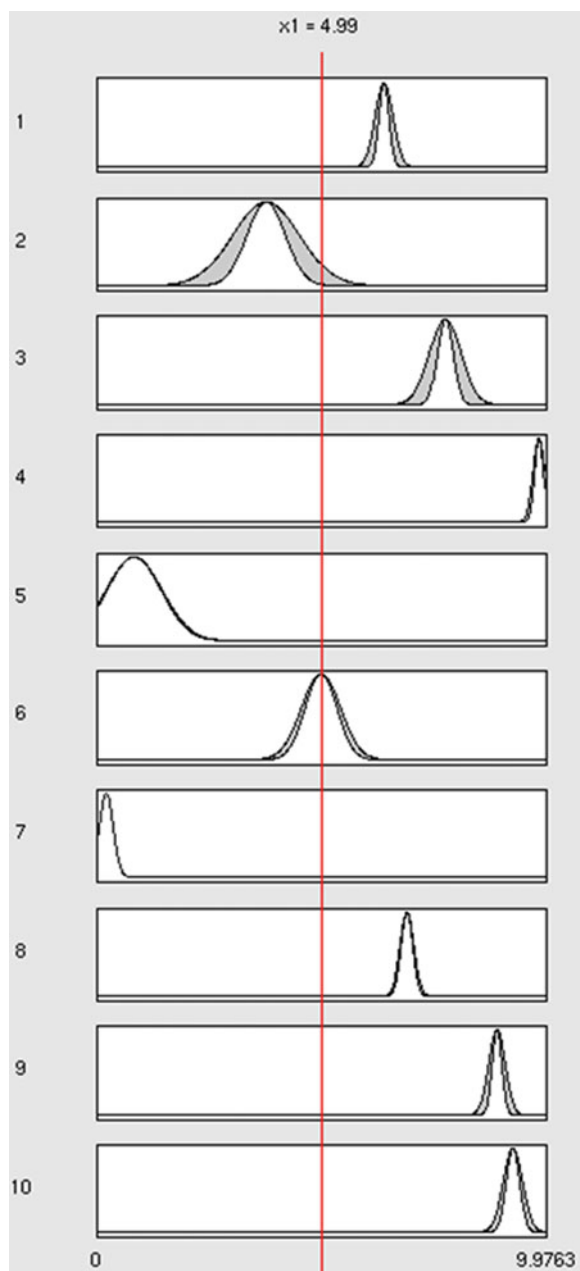
### 3.2 Results Discussion

The values obtained for the classification accuracy and RMSE error are not the best values obtained in general, yet they are comparable to current algorithms in terms of mean results [22–24]. This is by no manner the best obtainable results this approach can acquire; this is mostly in part to the chosen clustering algorithm as well as the evolutionary algorithm which were used to obtain such results. A better combination as well as tuning should yield better results.

As shown in the formed IT2 Fuzzy information granules, some granules captured more uncertainty than others, in many cases the uncertainty is minimal to the point that there is no measurable uncertainty when forming the evaluation of uncertainty Gaussian function.

Having chosen IT2 Fuzzy Gaussian membership functions as representation for higher type information granules, the characteristics of these is that the center value is the same, and only two values for  $\sigma$  form the FOU. Although results are acceptable, other variations can be used to yield different results as well as different interpretations, for example, where the center is offset and two values for  $\sigma$  are used. Even other types of IT2 Fuzzy membership functions could be used, each one having their own interpretation of the information as well as varying results when the IT2 Fuzzy System is formed and optimized.

**Fig. 6** Sample of the formed IT2 Fuzzy information granules for the 5th order polynomial dataset





## 4 Conclusion and Future Work

### 4.1 Conclusions

Taking inspiration from the uncertainty-based information theory, higher type information granules can be formed which better conceptualize the uncertainty in the information.

The proposed approach reduces the uncertainty in the information model by measuring the uncertainty by means of the difference between two evaluations of uncertainty created by two distinct measurements of information sampling.

By choosing Interval Type-2 Fuzzy sets as the representation of information granules, the proposed approach directly takes the obtained uncertainty measurement and builds higher type information granules.

Any other form of granule representation which can express the uncertainty in the information can be used [36].

### 4.2 Future Work

Find the optimal amount of samples for each model building step. Although 40 % was used, what is the minimal amount which can be used to obtain acceptable results?

The amount of samples taken could be explored; this chapter only took two samples to form the final information granule. Could taking more samples yield a better result?

Other information granule representations could be used which also support uncertainty. Even though Type A Gaussian evaluations of uncertainty were used, there are other types of functions which could also directly capture uncertainty.

**Acknowledgments** We thank the MyDCI program of the Division of Graduate Studies and Research, UABC, and Tijuana Institute of Technology the financial support provided by our sponsor CONACYT contract grant number: 314258.

## References

1. Bargiela, A., Pedrycz, W.: Toward a theory of granular computing for human-centered information processing. *IEEE Trans. Fuzzy Syst.* **16**(2), 320–330 (2008)
2. Pedrycz, W.: Granular computing: the emerging paradigm. *J. Uncertain Syst.* **1**(1), 38–61 (2007)
3. Pawlak, Z.: Rough sets. *Int. J. Comput. Inf. Sci.* **11**(5), 341–356 (1982)
4. Zhang, L.Z.L., Zhang, B.Z.B.: Quotient space based multi-granular computing, vol. 1 (2005)
5. Pedrycz, W., Vukovich, G.: Granular computing with shadowed sets. *Int. J. Intell. Syst.* **17**(2), 173–197 (2002)

6. Zhang, Y.Z.Y., Zhu, X.Z.X., Huang, Z.H.Z.: Fuzzy Sets Based Granular Logics for Granular Computing. (2009)
7. Yao, J.T.: Information granulation and granular relationships. In: 2005 IEEE International Conference on Granular Computing, vol. 1 (2005)
8. Yu, F.Y.F., Cai, R.C.R.: Optimized fuzzy information granulation of temporal data. In: Fuzzy Systems and Knowledge Discovery (FSKD), vol. 1 (2010)
9. Yao, J., Yao, Y.Y.: Information granulation for web-based information support systems. In: Proceedings of SPIE, pp. 138–146 (2003)
10. Park, H.-S., Chung, Y.-D., Oh, S.-K., Pedrycz, W., Kim, H.-K.: Design of information granule-oriented RBF neural networks and its application to power supply for high-field magnet. *Eng. Appl. Artif. Intell.* **24**(3), 543–554 (2011)
11. Sanchez, M.A., Castro, J.R., Perez-Ornelas, F., Castillo, O.: A hybrid method for IT2 TSK formation based on the principle of justifiable granularity and PSO for spread optimization. In: 2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), pp. 1268–1273 (2013)
12. Sanchez, M.A., Castro, J.R., Castillo, O.: Formation of general type-2 Gaussian membership functions based on the information granule numerical evidence. In: 2013 IEEE Workshop on Hybrid Intelligent Models and Applications (HIMA), pp. 1–6 (2013)
13. Klir, G.J.: *Uncertainty and Information: Foundations of Generalized Information Theory*, p. 499. Wiley-IEEE Press, New Jersey (2005)
14. Klir, G.J., Wierman, M.J.: *Uncertainty-Based Information*, vol. 15. Physica-Verlag HD, Heidelberg (1999)
15. Weise, K., Woger, W.: A Bayesian theory of measurement uncertainty. *Meas. Sci. Technol.* **3**, 1–11 (1992)
16. J. C. F. G. I. M. JCGM: Evaluation of measurement data—guide to the expression of uncertainty in measurement. *Int. Organ. Stand. Geneva ISBN.* **50**, 134 (2008)
17. Chiu, S.L.: Fuzzy model identification based on cluster estimation. *J. Intell. Fuzzy Syst.* **2**, 267–278 (1994)
18. Jang, J.-S.R.: Fuzzy modeling using generalized neural networks and Kalman filter algorithm. In: Proceedings of the Ninth National Conference on Artificial Intelligence, pp. 762–767 (1991)
19. Jang, J.S.R.: ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. man Cybern.* **23**(3), 665–685 (1993)
20. Yang, X.-S., Cuckoo search via Lévy flights. In: 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), pp. 210–214 (2009)
21. Frank A., Asuncion, A.: UCI Machine Learning Repository. University of California Irvine School of Information, School of Information and Computer Sciences, University of California, Irvine, p. 0 (2010) (vol. 2008, no. 14/8)
22. Daneshgar, A., Javadi, R., Razavi, B.S.: Clustering using isoperimetric number of trees. *Pattern Recognit.* **46**(12), 3371–3382 (2012)
23. David, G., Averbuch, A.: SpectralCAT: categorical spectral clustering of numerical and nominal data. *Pattern Recognit.* **45**(1), 416–433 (2012)
24. Thi, L., An, H., Hoai, L., Dinh, P.: New and efficient DCA based algorithms for minimum sum-of-squares clustering. *Pattern Recognit.* **47**, 388–401 (2014)
25. Castillo, O., Huesca, G., Valdez, F.: Evolutionary computing for topology optimization of type-2 fuzzy controllers. *Stud. Fuzziness Soft Comput.* **208**, 163–178 (2008)
26. Castillo, O., Melin, P.: *Type-2 Fuzzy Logic: Theory and Applications*. Springer-Verlag, Heidelberg (2008)
27. Castillo, O., Aguilar, L.T., Cazarez-Castro, N.R., Cardenas, S.: Systematic design of a stable type-2 fuzzy logic controller. *Appl. Soft Comput. J.* **8**, 1274–1279 (2008)
28. Castillo, O., Martinez-Marroquin, R., Melin, P., Valdez, F., Soria, J.: Comparative study of bio-inspired algorithms applied to the optimization of type-1 and type-2 fuzzy controllers for an autonomous mobile robot. *Inf. Sci.* **192**, 19–38 (2012)

29. Castillo, O., Melin, P., Alanis, A., Montiel, O., Sepulveda, R.: Optimization of interval type-2 fuzzy logic controllers using evolutionary algorithms. *J. Soft Comput.* **15**(6), 1145–1160 (2011)
30. Castro, J.R., Castillo, O., Melin, P.: An interval type-2 fuzzy logic toolbox for control applications. In: *Proceedings of FUZZ-IEEE 2007, London*, pp. 1–6 (2007)
31. Castro, J.R., Castillo, O., Martinez, L.G.: Interval type-2 fuzzy logic toolbox. *Eng. Lett.* **15**(1), 14 (2007)
32. Hidalgo, D., Castillo, O., Melin, P.: Type-1 and type-2 fuzzy inference systems as integration methods in modular neural networks for multimodal biometry and its optimization with genetic algorithms. *Inf. Sci.* **179**(13), 2123–2145 (2009)
33. Aguilar, L., Melin, P., Castillo, O.: Intelligent control of a stepping motor drive using a hybrid neuro-fuzzy ANFIS approach. *J. Appl. Soft Comput.* **3**(3), 209–219 (2003)
34. Melin, P., Castillo, O.: Adaptive intelligent control of aircraft systems with a hybrid approach combining neural networks, fuzzy logic and fractal theory. *J. Appl. Soft Comput.* **3**(4), 353–362 (2003)
35. Montiel, O., Sepulveda, R., Melin, P., Castillo, O., Porta Garcia, M., Meza Sanchez I.: Performance of a simple tuned fuzzy controller and a PID controller on a DC motor. In: *FOCI 2007* pp. 531–537
36. Rubio, E., Castillo, O.: Optimization of the interval type-2 fuzzy C-means using particle swarm optimization. In: *Proceedings of NABIC 2013*, pp. 10–15, Fargo, USA (2013)