Layered URC fuzzy systems: a novel link between fuzzy systems and neural networks

Jeffrey J. Weinschenk[a], Robert J. Marks II[a], William E. Combs[b]

[a]Department of Electrical Engineering
University of Washington
253 EE/CSE Building
Campus Box 352500
Seattle, WA 98195-2500
jjw77@ee.washington.edu

[b]The Boeing Company
P.O. Box 3707
MS: 6F-13
Seattle, WA 98124
william.e.combs@boeing.com

Abstract— **We introduce a novel layered fuzzy architecture that avoids rule explosion. Unlike a single layer union rule configuration (URC) fuzzy system, a layered URC fuzzy system can approximate any surface without the need of burdensome "corrective" terms. Further, we show that the URC fuzzy system is a generalized layered perceptron—an insight that allows one to choose interconnection weights in an intuitive manner with very basic problem knowledge. In some cases, training may not be necessary. Further, the fuzzy linguistic meaning of variables is preserved throughout the layers of the system. The universal approximation property of this architecture is discussed and we demonstrate how a layered URC fuzzy system solves a simple regression problem.**

Index Terms—**union rule configuration, fuzzy logic, rule explosion, universal approximation, layered perceptron**

I. INTRODUCTION

For many years researchers have developed neural networks and fuzzy systems for a wide range of applications. Recently, much effort has been focused on neuro-fuzzy systems in an attempt to activate the desirable properties of both types of systems in a single network [1-4].

We present a novel layered union rule configuration (URC) fuzzy system architecture that bridges an important gap between fuzzy systems and neural networks. This layered fuzzy architecture is a generalized form of the layered perceptron and encompasses virtually all feed-forward architectures including those that incorporate sigmoids, radial basis functions, or other kernels. By viewing the layered URC fuzzy system as a universal architecture, knowledge of fuzzy systems and layered perceptrons combine to give each weight a significant and intuitive meaning.

Section 2 contains of a brief review of URC fuzzy systems. A discussion of the layered URC fuzzy system architecture follows in Section 3. The necessary number of layers, as well as the purpose and construction of each layer, is discussed. Further, the fuzzy linguistic meaning of intermediate, inter-layer, variables is examined. A discussion on the universal approximation property of the layered URC fuzzy system is also included. Practical implementation

issues are discussed in Section 4 as well as a simple regression example. Concluding remarks follow in Section 5.

II. A REVIEW OF THE UNION RULE CONFIGURATION

Conventional fuzzy systems are composed of multi-antecedent rules connected by the intersection operator. The output formula for this type of system, with sum product logic and multidimensional rule table **F**, is given by

$$z_{IRC}(\mathbf{x}) = \frac{\sum_{a_1=1}^{N_1}\sum_{a_2=1}^{N_2}\cdots\sum_{a_P=1}^{N_P}\mathbf{F}(a_1,a_2,\cdots,a_P)\prod_{j=1}^{P}\mu_{j,a_j}(x_j)}{\sum_{a_1=1}^{N_1}\sum_{a_2=1}^{N_2}\cdots\sum_{a_P=1}^{N_P}\prod_{j=1}^{P}\mu_{j,a_j}(x_j)}, \quad (1)$$

where the value of membership of the input x_j to the a_jth subset of the jth antecedent is $\mu_{j,a_j}(x_j)$, N_j is the number of subsets of the jth antecedent, and the a_j's index **F**. Notice that when the number of antecedents is increased, this system is burdened by an exponential increase in the number of fuzzy rules, giving rise to *rule explosion*.

In contrast, W. E. Combs *et al.* propose a novel fuzzy system architecture where the union operator connects single antecedent rules [5]. A recent paper by Weinschenk *et al.* [6] formalizes the output formula for this type of system for sum-product logic and centroid defuzzification as

$$y_{URC}(\mathbf{x}) = \frac{\sum_{i=1}^{P}\sum_{j=1}^{N_i}y_{i,j}\mu_{i,j}(x_i)}{\sum_{i=1}^{P}\sum_{j=1}^{N_i}\mu_{i,j}(x_i)}, \quad (2)$$

where $y_{i,j}$ is a consequent center of mass and $\mu_{i,j}(x_i)$ is the value of membership of the input, x_i, to the jth subset of

the *i*th antecedent. This new architecture has the property that a linear increase in the number of antecedents results in a linear increase in the number of fuzzy rules—hence rule explosion is averted. W. E. Combs *et al.* refer to this novel architecture as a *union rule configuration* (URC) while the conventional fuzzy architecture is referred to as an *intersection rule configuration* (IRC).

In [6], Weinschenk *et al.* explore the relationship between IRC and URC fuzzy systems and prove that IRC and URC fuzzy systems are identical if three constraints are met. First, the IRC rule table must be *additively separable*. A multi-dimensional rule table **F** is said to be additively separable if

$$\mathbf{F} = \bar{v}_1 \oplus \bar{v}_2 \oplus \cdots \oplus \bar{v}_P, \tag{3}$$

where $\bar{v}_i$ is a *projection vector* corresponding to the *i*th antecedent of a multi-dimensional IRC rule table **F** and $\oplus$ is the outer sum operator.[1] Similarly, a function *g* is additively separable if

$$g(x_1, x_2, \cdots, x_P) = \sum_{i=1}^{P} g_i(x_i). \tag{4}$$

It is our experience that many IRC rule tables used in industry meet the criterion given in (3).

Secondly, the antecedent membership functions must be constructed such that

$$\sum_{k=1}^{N_i} \mu_{i,k}(x_i) = 1 \quad \forall \ i. \tag{5}$$

Although this second requirement is necessary in order to achieve strict equality between IRC and URC fuzzy systems, in practice it is often unnecessary and even undesirable. In an IRC fuzzy system, if a fuzzified input is reduced to noise the system reacts poorly because each antecedent is weighted equally. In the corresponding URC fuzzy system, if (5) is not enforced, the URC fuzzy system acts robustly to missing inputs by discounting low magnitude antecedents.

Finally, equivalence between IRC and URC fuzzy systems requires that

$$y_{i,j} = Pv_i(j) \quad \forall \ (i,j). \tag{6}$$

Thus, elements of the projection vectors are related to elements of the URC rule table by a scaling constant, *P*.

Notice that substitution of (6) into (2) yields

---

[1] An outer sum is similar in nature to an outer product in that the outer sum of a set of *P* vectors results in a P dimensional matrix. The key difference is that an outer sum forms the P dimensional matrix via sums of elements whereas the outer product uses multiplication as the constructor.

$$y_{URC}(\mathbf{x}) = \sum_{i=1}^{P} \sum_{j=1}^{N_i} v_i(j) \frac{\mu_{i,j}(x_i)}{K}, \tag{7}$$

where

$$K = \frac{1}{P} \sum_{i=1}^{P} \sum_{j=1}^{N_i} \mu_{i,j}(x_i). \tag{8}$$

Thus, if one wants to retain the desirable property of the URC where low magnitude inputs are discounted, the constant *K* may be thought of as a scale factor that is either applied to the projection vectors or to the antecedent membership functions. Notice that when (5) is enforced, *K* is unity and strict equality between the IRC and URC exists.

When the desired IRC rule table is additively inseparable, equality between IRC and URC fuzzy systems is still achievable by adding corrective terms to the URC fuzzy system. Corrective terms are point masses placed on the output axis that counter-balance regions in the domain space inaccurately represented by a separable IRC rule table. Corrective terms follow in IRC structure, each corrective term corresponding to a multi-antecedent rule. One corrective term is required for each element of the IRC rule table that is inaccurately represented by the separable IRC rule table implemented by the URC fuzzy system.

It is well known that an IRC fuzzy system is a universal approximator [7-9]. Due to the equality of IRC and URC fuzzy systems, it follows that a URC fuzzy system, with corrective terms where necessary, is a universal approximator as well. However, depending on the nature of the additively inseparable IRC rule table, so many corrective terms may be required that the URC fuzzy system no longer offers significant computational savings over the analogous IRC fuzzy system. Thus, in the next section we present a layered URC architecture that maintains the universal approximation property in the absence of corrective terms.

III. LAYERED URC FUZZY SYSTEMS

A. A novel link between fuzzy systems and neural nets

In the interest of computational efficiency, it is advisable to avoid adding corrective terms to a URC fuzzy system. However, it is important to maintain the universal approximation property. With these goals in mind, we examine a layered URC fuzzy system without corrective terms. Fig. 1 depicts a 2-layer URC fuzzy system with three subsets per antecedent, where the outputs of one URC fuzzy layer serve as the inputs to the next. Arrows represent

weighted interconnects, where the interconnect weights are simply elements of the projection vectors as specified in (7).
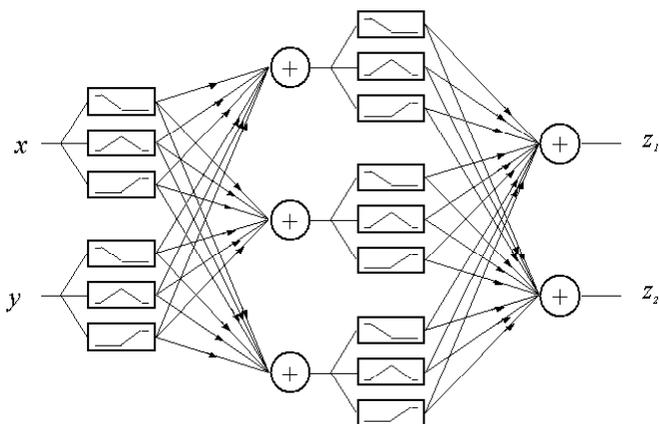

Fig. 1. A 2-layer URC fuzzy system with three subsets per antecedent.

Notice that the 2-layer URC fuzzy system in Fig. 1 appears similar in nature to a feed-forward neural network. The only major difference is that the output of each node is fed into a set of membership functions instead of a common sigmoid or other kernel function. Fig. 2 shows another URC fuzzy system, with only one subset per antecedent, where a common sigmoid replaces all of the membership functions. The first set of sigmoids make up an input layer, the three summations followed by sigmoids make up a single hidden layer, and the final two summations compose an output layer. It is apparent that the layered URC fuzzy system (without corrective terms) is a generalized feed-forward neural network. Notice that the use of sigmoids, radial basis functions, and other kernels common to neural networks can be readily interpreted as different types of membership functions.
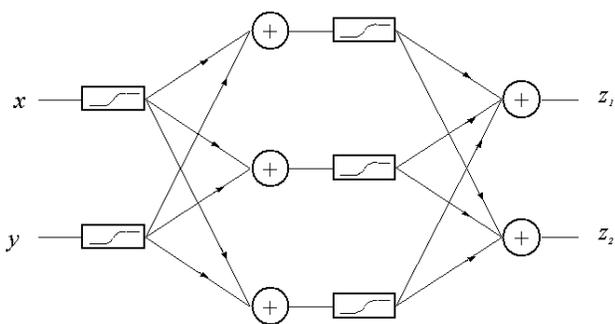

Fig. 2. A 2-layer URC fuzzy system with one subset per antecedent and the common sigmoid for all membership functions.

An interesting parallel exists between the layered URC and the layered perceptron. Ever since the classic objection put forth by Minsky and Papert [10], it was known that a single-layer perceptron (no hidden layer) is not a universal approximator. Similarly, a single-layer URC fuzzy system without corrective terms is also not a universal approximator.

However, since the layered perceptron's architecture is subsumed by the 2-layer URC fuzzy system, and a 2-layer perceptron (containing one hidden layer) is known to be a universal approximator [11], a 2-layer URC fuzzy system must also be a universal approximator. Importantly, this means a 2-layer URC architecture maintains the universal approximation capacity in the absence of corrective terms.

B. Rotations of Fuzzy Variables

Since a 2-layer URC fuzzy system without corrective terms is a universal approximator, it follows that a 2-layer URC fuzzy system can perform the same function as an IRC fuzzy system that has an additively inseparable rule table. As discussed previously, each individual URC layer can only solve problems involving an additively separable IRC rule table. Therefore the first URC layer must take inputs, which relate to the desired output through an additively inseparable rule table, and convert them (via an additively separable IRC rule table) to a set of intermediate variables that relate to the output through an additively separable rule table. To avoid confusion, from this point on we refer to the desired (potentially additively inseparable) IRC rule table as the *primary rule table*. Further, the primary rule table relates the *primary inputs* to the *primary outputs*. The additively separable IRC rule tables that govern the behavior of the first and second URC layers are referred to as *first and second layer rule tables*, respectively. For the following discussion, it is useful to think of the primary rule table as containing samples from a multi-dimensional surface where each element of the primary rule table is the center of mass of a consequent subset.

Fundamental tomography dictates that any surface can be represented to an arbitrary accuracy via a superposition of projections where, by definition, the outer sum of orthogonal projections forms an additively separable matrix [12]. Thus, the primary rule table is expressible as a superposition of additively separable surfaces, each of which is formed from a different pair of orthogonal projections that occur at different angles relative to the primary inputs. Importantly, each of these additively separable surfaces may be implemented with a URC fuzzy system that relates rotated primary inputs to an intermediate output. However, since the terminal operation of a URC layer is summation, superimposing the intermediate outputs of a set of URC fuzzy systems is equivalent to constructing one large composite URC layer that relates all of the rotated primary inputs directly to the primary output. This composite URC fuzzy system effectively implements the second layer rule table and constitutes the second layer of a 2-layer URC fuzzy system.

The first layer of the 2-layer URC fuzzy system must effectively rotate the primary inputs to the necessary angles. A rotation in two dimensions is given by

$$r_\theta = x \cos(\theta) + y \sin(\theta), \qquad (9)$$

where θ is the angle of rotation.[2] A URC fuzzy system implements a desired rotation by selecting its additively separable rule table to be samples from the surface defined by (9).[3] Thus, the first layer rule table is carefully chosen such that the necessary rotated primary inputs are generated. In general, the optimum rotation angles can be determined by standard search techniques [14].

Interestingly, rotated fuzzy variables do retain fuzzy linguistic meaning. In the case of $\pm\pi/4$ radian rotations, the rotated fuzzy variables give respective measures of the midpoint and separation between the membership functions of the two antecedents. In general, rotations through other angles yield a weighted measure of the midpoint or separation between membership functions.

The layered URC provides a maximum computational savings when the primary rule table is accurately represented by a small number of projections. In the worst case, when the primary rule table contains random numbers, the layered URC fuzzy system will actually require more resources than the analogous IRC fuzzy system. However, it is our experience that many additively inseparable IRC rule tables found in the literature can be accurately characterized with only a few projections.

IV. IMPLEMENTATION ISSUES AND A REGRESSION EXAMPLE

To gain a better understanding of the behavior of 2-layer URC fuzzy systems, we examine a toy problem with an inseparable rule table—the fuzzy XOR. A fuzzy XOR is defined as a function $z(x, y)$ on $x, y \in [0,1]$ that satisfies the boundary conditions

$$z(0,0) = z(1,1) = 0 \qquad (10)$$

and

$$z(0,1) = z(1,0) = 1 \qquad (11)$$

[15]. One such function is given by

$$z_1(x, y) = 0.5 - 2(x - 0.5)(y - 0.5). \qquad (12)$$

An IRC fuzzy system can be constructed that approximates (12) by choosing a set of ordered membership functions for both inputs and then forming a rule table from samples of the

surface defined by (12).[4] However, since (12) is not additively separable as defined by (4), the resulting IRC rule table will be not be additively separable. Thus, a single-layer URC fuzzy system requires corrective terms for accurate approximation of (12) and, in this case, so many corrective terms are required that the single-layer URC fuzzy system is more computationally expensive than the analogous IRC fuzzy system.

However, if (12) is rotated by $\pi/4$ radians about the point $(0.5, 0.5)$, one obtains the new surface,

$$z_2(x, y) = 0.5 + 2(x - 0.5)^2 - 2(y - 0.5)^2. \qquad (13)$$

This rotated fuzzy XOR surface is additively separable as defined by (4), thus an IRC rule table formed from its samples will also be additively separable. It is therefore apparent that a 2-layer URC fuzzy system can approximate (12) if the first layer rotates the inputs by $\pm\pi/4$ radians and the second layer implements the separable rule table formed from samples of (13). In general, more complex surfaces require more than one pair of orthogonal projections.

The first layer of the 2-layer URC system is constructed such that each antecedent has three subsets: LOW, MED, and HIGH. The first layer has two consequents, the first of which represents a rotation of the inputs by $\pi/4$ radians while the second represents a rotation of the inputs by $-\pi/4$ radians. Recall that an additively separable IRC rule table that performs these rotations is found by sampling (9). The URC interconnect weights are then identified from the additively separable IRC rule table via (3) and (7). To ensure that the rotated inputs $r_{\pi/4}$ and $r_{-\pi/4}$ remain on $[0,1]$, it is necessary to shift and scale the surface in (9) such that no consequent subset has a center of mass outside $[0,1]$. The URC interconnect weights for the first layer are given in Table 1. The antecedent membership functions are shown in Fig. 3.
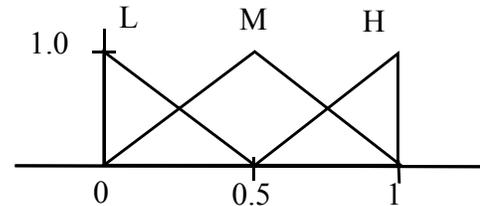

Fig. 3. The antecedent membership functions for the first layer of the 2-layer URC fuzzy system. Both antecedents use the same set of membership functions. Notice that the membership functions sum to unity.

The surface in (13) is sampled in order to obtain the second layer rule table. The URC interconnect weights are

---

[2] In the interest of clarity we discuss a rotation in 2D although in general, rotations can be performed in higher dimensional spaces.

[3] We point out that rotations need not be implemented with a URC layer. Instead, the rotations can be performed on normalized inputs in a preprocessing step, in which case a single layer URC fuzzy system, with no corrective terms, is sufficient for universal approximation. This method is reminiscent of Yoh-Han Pao's functional link neural network [13].

[4] Each sample represents the center of mass of a consequent subset.

obtained from the second layer rule table via (3) and (7). Table 2 contains interconnect weights for the case where each antecedent has 3 subsets. The antecedent membership functions for the second layer are the same as those used in the first layer.

	x			y		
	LOW	MED	HIGH	LOW	MED	HIGH
$r_{\pi/4}$	0	0.25	0.5	0	0.25	0.5
$r_{-\pi/4}$	0	0.25	0.5	0.5	0.25	0

Table 1. The interconnect weights for the first layer of the 2-layer URC fuzzy system. These weights implement a +/- $\pi/4$ degree rotation of the inputs.

$r_{\pi/4}$			$r_{-\pi/4}$		
LOW	MED	HIGH	LOW	MED	HIGH
0	0.5	0	0.5	0	0.5

Table 2. The interconnect weights for the second layer of the 2-layer URC fuzzy system.

Fig. 4 depicts the output surface for the 2-layer URC fuzzy system for the case where the second layer antecedents have three subsets each. For the purpose of comparison, the output surface for the IRC fuzzy system that directly approximates (12) is shown in Fig. 5. The IRC fuzzy system implements an inseparable 3x3 rule table (Table 3) obtained by sampling (12). The IRC fuzzy system uses the same triangular membership functions depicted in Fig. 3.
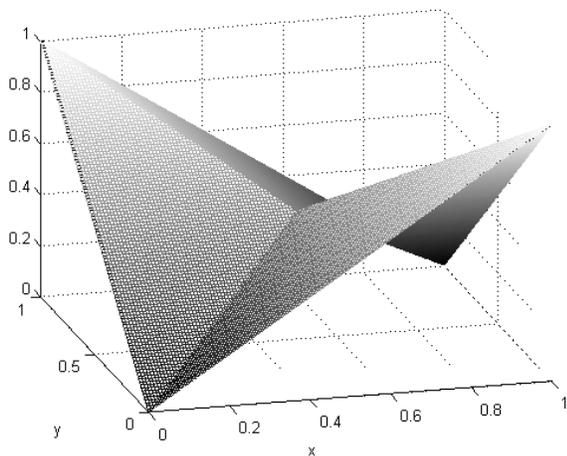

Fig. 4. The 2-layer URC output surface. Each antecedent in the second layer has three subsets each.

Notice that the output surfaces shown in Figures 4 and 5 have the same general shape, yet are not identical. This is an interesting phenomenon whose explanation lies in the inherent structure of the URC fuzzy system. The IRC fuzzy system forms an output surface by linearly interpolating between samples from the original surface given in (12). In contrast, the URC fuzzy system linearly interpolates between samples from the rotated surface given in (13) and then rotates the result back to the proper orientation. Thus, both output surfaces are legitimate, yet different, approximations of (12). In fact, had the desired surface been that shown in Fig. 4 and not that of (12), the corresponding IRC and URC fuzzy systems would still yield the same output surfaces shown in Figures 4 and 5, yet it would be the IRC fuzzy system that is less accurate!
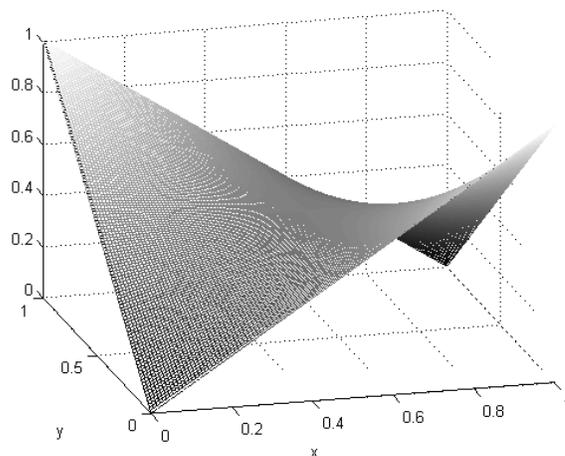

Fig. 5. The IRC output surface. Each antecedent has three subsets.

		x		
		LOW	MED	HIGH
	LOW	0	0.5	1
y	MED	0.5	0.5	0.5
	HIGH	1	0.5	0

Table 3. The additively inseparable IRC rule table used to directly approximate (12).

It is important to note that the accuracy of both the IRC and URC approximations can be improved by increasing the number of antecedent subsets. However, since IRC fuzzy systems are crippled by rule explosion, the addition of antecedent subsets to an IRC fuzzy system can be costly. Alternatively, antecedents are easily added to the second layer of the 2-layer URC fuzzy system to improve approximation accuracy. In fact, each new additional antecedent subset results in one new interconnect weight. Figures 6 and 7 depict approximations for the cases where the 2-layer URC fuzzy system has 5 and 9 subsets per antecedent in the second layer, respectively. Notice that as more antecedent subsets are added to the second layer, the

accuracy of the approximation improves. Table 4 shows the maximum absolute error (XAE) and the mean absolute error (MAE) for the three URC approximations.
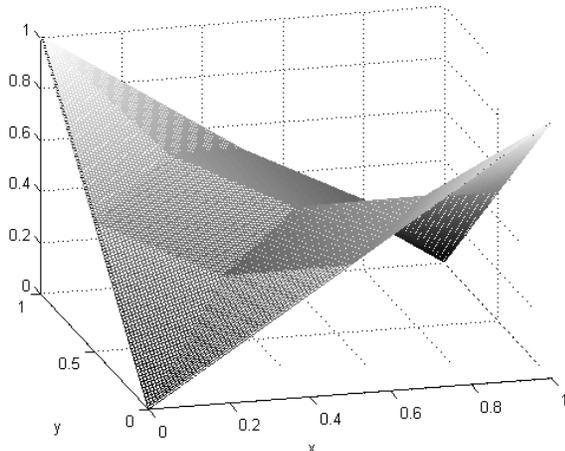

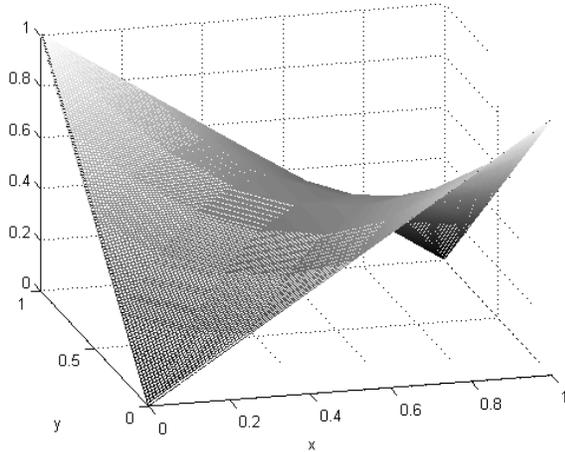Fig. 6. The 2-layer URC output surface. Each antecedent in the second layer has five subsets each.

Fig. 7. The 2-layer URC output surface. Each antecedent in the second layer has nine subsets each.

	# of subsets / antecedent in the second layer		
	3	5	9
XAE	0.1250	0.0312	0.0079
MAE	0.0410	0.0102	0.0026

Table 4. The XAE and MAE for the three URC approximations. Notice the error quickly diminishes as the number of subsets per second-layer antecedent increases.

V. CONCLUSION

Many IRC fuzzy systems have additively separable rule tables, in which case a single-layer URC can be used to avoid rule explosion and provide the exact same functionality as the original IRC fuzzy system. However, for those IRC fuzzy systems with additively inseparable rule tables, rule explosion can still be avoided in most instances through the use of a 2-layer URC architecture. Although a 2-layer URC fuzzy system yields a slightly different output surface than the corresponding IRC fuzzy system, both output surfaces are valid representations of the IRC rule table. Indeed, as the number of antecedent subsets increases, the difference between the 2-layer URC output surface and the IRC output surface diminishes to zero. Finally, we invite the reader to design a URC fuzzy system and find out first hand how rule explosion is avoided.

References

[1] S. V. Kartalopoulos, *Understanding Neural Networks and Fuzzy Logic : Basic Concepts and Applications*. Wiley-IEEE Press, 1995.
[2] H.-X. Li, C. L. P. Chen, H.-P. Huang, H. Li, Fuzzy Neural Intelligent Systems: Mathematical Foundation and the Applications in Engineering. CRC Press, 2000.
[3] D. Nauck, R. Kruse, F. Klawonn, Foundations of Neuro-Fuzzy Systems. John Wiley & Sons, 1997.
[4] J. Buckley, Y. Hayashi, E. Czogaia, "On the equivalence of neural networks and fuzzy expert systems," *IJCNN*, vol. 2 pp. 7-11, Jun. 1992.
[5] W. E. Combs and J. E. Andrews, "Combinatorial rule explosion eliminated by a fuzzy rule configuration," *IEEE Trans. Fuzzy Systems*, vol. 6, no. 1, pp. 1-11, Feb. 1998.
[6] J. J. Weinschenk, W. E. Combs, R. J. Marks II, "Avoidance of rule explosion by mapping fuzzy systems to a disjunctive rule configuration," *FUZZIEEE*, accepted for publication Jan. 2003.
[7] L.-X. Wang and J. M. Mendel, "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning," *IEEE Trans. Neural Networks*, vol. 3, no. 5, pp. 807-814, Sept. 1992.
[8] L.-X. Wang, "Fuzzy systems are universal approximators," *Proc. IEEE Int'l. Conference on Fuzzy Systems*, San Diego, CA 1992, pp. 1163-1170.
[9] V. Kreinovich, G. C. Mouzouris and H. T. Nguyen, "Fuzzy rule based modeling as a universal approximation tool," *Fuzzy Systems, Modeling and Control* (H. T. Nguyen and M. Sugeno, Eds.), pp. 135-195, Kluwer Ac. Publ., Boston, 1998.
[10] M. L. Minsky, S. A. Papert, *Perceptrons - Expanded Edition: An Introduction to Computational Geometry*. MIT Press, 1987.
[11] G. Cybenko, "Apprximations by superpositions of a sigmoidal function," *Math. Control, Signals and Systems* vol. 2, pp. 303-314, 1989.
[12] D. E. Dudgeon, R. M. Mersereau, *Multidimensional Digital Signal Processing*. Prentice Hall, 1984.
[13] Y.-H. Pao, *Adaptive Pattern Recognition and Neural Networks*. Addison-Wesley Pub Co., 1989.
[14] R. D. Reed, R. J. Marks II, *Neural Smithing*. The MIT Press, Cambridge MA, 1998.
[15] G. J. Klir, T. A. Folger, C. Papa, *Fuzzy Sets, Uncertainty and Information*. Prentice Hall, 1988.