Avoidance of rule explosion by mapping fuzzy systems to a union rule configuration

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

[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 present a novel mapping whereby a classical fuzzy system, based on an intersection rule configuration (IRC), is converted to a union rule configuration (URC) system. Previous work has demonstrated that URC fuzzy systems avoid rule explosion, where a linear increase in the number of antecedents gives rise to an exponential increase in the number of fuzzy rules. However, there has been some doubt as to the validity of URC systems and previous findings. We resolve lingering questions and prove that any arbitrary IRC system can be converted to a URC system with identical performance. Further, we show that URC systems do avoid rule explosion for many problems. Finally, we note that a URC system is a universal approximator.**

***Index Terms*—fuzzy logic, union rule configuration, intersection rule configuration, rule explosion, universal approximation.**

I. Introduction

CURRENTLY, most fuzzy logic systems are based on an intersection rule configuration (IRC) where rules map antecedent subsets, connected with the intersection operator, to a consequent subset. These multi-antecedent rules incorporate an expert's perceived correlation between antecedents. However, as antecedents are added to an IRC system, the number of fuzzy rules increases exponentially, giving rise to *rule explosion.*

Combs and Andrews suggest an alternative rule configuration that avoids rule explosion [1]. Their method is *inspired* by the propositional logic expression

$$[(p \cap q) \Rightarrow r] \Leftrightarrow [(p \Rightarrow r) \cup (q \Rightarrow r)], \qquad (1)$$

where p and q are antecedents, r is a consequent, $\cap$ represents intersection, $\cup$ represents union, and $\Rightarrow$ represents the implication operator. In this alternative rule configuration, multi-antecedent rules are transformed into single-antecedent rules. Single antecedent rules that have the same antecedent are combined. Thus a linear increase in the number of antecedents results in a linear increase in the number of rules hence rule explosion is averted. Combs and Andrews refer to this single antecedent architecture as a union rule configuration (URC).

Some have questioned the validity of the URC since it was inspired by the propositional relation in (1) which is not true in general for fuzzy logic [2-5]. There is also some question as to whether single antecedent questions can really replace multi-antecedent questions while still retaining an expert's perceived correlation between antecedents. Further, since a URC fuzzy system contains fewer *degrees of freedom*[1] than an IRC fuzzy system, how is it possible to find equality between the two? In this paper we address the above-mentioned issues as we prove that equality does exist between IRC and URC systems. Further, we demonstrate how to convert any arbitrary IRC system into a URC system. As a consequence of the equality between IRC and URC systems, we will note that URC systems must be universal approximators.

In Section II a class of additively separable IRC rule tables are defined that allow for a direct mapping from the IRC to the URC architecture. Section II closes with a discussion on how to map inseparable IRC rule tables to URC rule tables. In Section III an example is presented where a traditional IRC rule table is mapped to a URC rule table. Finally, a few concluding remarks are offered in Section IV.

II. A Mapping From the IRC to the URC

First, a mapping from IRC to URC fuzzy systems is presented for a special class of IRC rule tables. Many IRC rule tables used in the industry fit into this class. IRC rule tables that are not strictly contained within this class require an additional step in the conversion process and are the topic of Subsection B.

[1] By this term we refer to the extent of control a system has over the problem space. The IRC grants control over all possible combinations of antecedent subset intersections, while the URC grants control over individual antecedent subsets.

0-7803-7810-5/03/$17.00 ©2003 IEEE

A. A Mapping for Additively Separable IRC Rule Tables

Consider an IRC rule table that relates P antecedents to a single consequent where each antecedent contains a distinct number of fuzzy subsets.[2] Let the system utilize centroid defuzzification where consequent subsets may take on any desired shape. It is useful to think of a consequent subset in terms of its center of mass. Thus, an IRC rule table simply maps weights, determined by an intersection of antecedent subsets, to locations specified by the centers of mass of the consequent subsets.

Let $\bar{v}_i$ be a *projection vector* corresponding to the ith antecedent of a multi-dimensional IRC rule table $\mathbf{F}$ such that

$$\mathbf{F} = \bar{v}_1 \oplus \bar{v}_2 \oplus \cdots \oplus \bar{v}_P, \tag{2}$$

where $\oplus$ is the outer sum operator[3]. We denote any IRC rule table that satisfies (2) *additively separable.* A consequence of additive separability is

$$\mathbf{F}(a_1, a_2, \cdots, a_P) = \sum_{i=1}^{P} v_i(a_i), \tag{3}$$

where the a_i's index $\mathbf{F}$. In other words, any element in the rule table is expressible as a sum of projection elements with each projection vector contributing one element. The set of projection vectors $[\bar{v}_1, \bar{v}_2, \cdots, \bar{v}_P]$ is not unique, however, as the set $[\bar{v}_1 + c_1, \bar{v}_2 + c_2, \cdots, \bar{v}_P + c_P]$ will yield the same rule table provided that constants $c_1, c_2, \cdots, c_P$ sum to zero.

Additive separability, while a strong constraint, is not necessarily inconvenient in many design problems. Many rule tables monotonically increase from one corner to the opposite corner. These tables are often additively separable. In some cases, the corners of these rule tables exhibit an artificial, 'saturated' behavior due a designer's effort to reduce the number of consequent subsets. These saturated regions often ruin the additive separability of the rule table in exchange for a greater dependence on tuning[4]. Of course rule tables need not be monotonic in all of the antecedents to be additively separable.

Equality between IRC and URC systems is first established for additively separable IRC fuzzy systems. The output formula for a URC system with sum-product logic and centroid defuzzification is given by

$$y_{URC}(\mathbf{x}) = \frac{\sum_{i=1}^{P}\sum_{j=1}^{N_i} y_{i,j}\mu_{i,j}(x_i)}{\sum_{i=1}^{P}\sum_{j=1}^{N_i} \mu_{i,j}(x_i)}, \tag{4}$$

where the centers of mass of the consequent subsets are specified by $y_{i,j}$, the value of membership of the input x_i to the jth subset of the ith antecedent is $\mu_{i,j}(x_i)$, and N_i is the number of subsets of the ith antecedent. This expression differs only slightly from that found in [6] (1) and [7] (30) in that it allows each antecedent to have a distinct number of subsets. From this point on, 'x_i' in the notation for the membership functions will be dropped.

In contrast, the output formula for an IRC fuzzy system with sum product logic and multidimensional rule table $\mathbf{F}$ is given as

$$z_{IRC}(\mathbf{x}) = \frac{\sum_{a_1=1}^{N_1}\sum_{a_2=1}^{N_2}\cdots\sum_{a_P=1}^{N_P} \mathbf{F}(a_1, a_2, \cdots, a_P)\prod_{j=1}^{P}\mu_{j,a_j}}{\sum_{a_1=1}^{N_1}\sum_{a_2=1}^{N_2}\cdots\sum_{a_P=1}^{N_P}\prod_{j=1}^{P}\mu_{j,a_j}}, \tag{5}$$

where once again N_i is the number of subsets of the ith antecedent and the a_i's index $\mathbf{F}$. If the IRC rule table is additively separable, substitution of (3) results in

$$z_{IRC}(\mathbf{x}) = \frac{\sum_{i=1}^{P}\sum_{a_1=1}^{N_1}\sum_{a_2=1}^{N_2}\cdots\sum_{a_P=1}^{N_P} v_i(a_i)\prod_{j=1}^{P}\mu_{j,a_j}}{\prod_{j=1}^{P}\left(\sum_{k=1}^{N_j}\mu_{j,k}\right)}, \tag{6}$$

where the denominator of (5), a summation over an outer product of vectors of antecedent membership values, is factored into a product of sums.

In (6), numerator absorbs the denominator to yield

[2] This discussion easily generalizes to multiple consequent systems, but in the interest of simplifying the explanation we restrict ourselves to single consequent systems.

[3] 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.

[4] It is our experience that fuzzy systems based on saturated rule tables often require significant tuning. In general, additively separable rule tables seem to rely less on tuning.

$$z_{IRC}(\mathbf{x}) = \sum_{i=1}^{P}\sum_{a_1=1}^{N_1}\sum_{a_2=1}^{N_2}\cdots\sum_{a_P=1}^{N_P} v_i(a_i)\prod_{j=1}^{P}\frac{\mu_{j,a_j}}{\sum_{k=1}^{N_j}\mu_{j,k}}. \quad (7)$$

Next, the term corresponding to $j = i$ in the product of (7) is pulled out to achieve

$$z_{IRC}(\mathbf{x}) = \sum_{i=1}^{P}\sum_{a_i=1}^{N_i}\frac{v_i(a_i)\mu_{i,a_i}}{\sum_{k=1}^{N_i}\mu_{i,k}}C, \quad (8)$$

where C is a constant given by

$$C = \sum_{a_1=1}^{N_1}\cdots\sum_{a_{i-1}=1}^{N_{i-1}}\sum_{a_{i+1}=1}^{N_{i+1}}\cdots\sum_{a_P=1}^{N_P}\prod_{\substack{j=1\\ j\neq i}}^{P}\frac{\mu_{j,a_j}}{\sum_{k=1}^{N_j}\mu_{j,k}} = 1. \quad (9)$$

Finally, observe that

$$P = \sum_{i=1}^{P}\sum_{j=1}^{N_i}\frac{\mu_{i,j}}{\sum_{k=1}^{N_i}\mu_{i,k}}, \quad (10)$$

since the sum over a set of normalized membership values is unity. Thus, (8) is multiplied by unity to achieve

$$z_{IRC}(\mathbf{x}) = \frac{\sum_{i=1}^{P}\sum_{j=1}^{N_i} P v_i(j)\frac{\mu_{i,j}}{\sum_{k=1}^{N_i}\mu_{i,k}}}{\sum_{i=1}^{P}\sum_{j=1}^{N_i}\frac{\mu_{i,j}}{\sum_{k=1}^{N_i}\mu_{i,k}}}. \quad (11)$$

Upon comparing (11) to (4), some interesting relationships become apparent. First, the consequent centers of mass of the URC fuzzy system are related to the projection vectors of the IRC rule table such that

$$y_{i,j} = P v_i(j) \quad \forall \ (i,j). \quad (12)$$

Thus, the consequent centers of mass of the URC system are scaled versions of the IRC projection vector elements. Therefore, in general, the mapping will hold for *any* method of defuzzification provided the consequent subsets are constrained to have the same shape and are multiplied by an appropriate scale factor. Of course, the consequent membership functions need not have identical shape if centroid defuzzification is employed (or any other defuzzification technique that depends only on the center of mass). No restrictions are placed on the shape of the antecedent membership functions.

Secondly, for equality between IRC and URC systems to hold, the URC system must be supplied with a set of normalized membership values for each antecedent, such that the membership values for a given antecedent sum to unity. Examination of (5) reveals that normalization of the membership values does not affect the output formula of the IRC, however, normalization of the membership values does affect the behavior of the URC system. URC systems, unlike the IRC systems, give greater weight to those antecedents whose vector of membership values has a large magnitude. Thus, if a fuzzified input were to drop out or be reduced to low magnitude noise, a URC system reacts robustly by discounting that antecedent, whereas the IRC system will weight all antecedents equally and perform poorly. Therefore although membership values must be normalized to obtain equality between the IRC and URC systems, it may worthwhile to skip this normalization step.

This mapping is an exciting result, as the IRC architecture requires

$$r_{IRC} = \prod_{i=1}^{P} N_i \quad (13)$$

rules while the URC architecture accomplishes the same task with

$$r_{URC} = \sum_{i=1}^{P} N_i \quad (14)$$

rules.[5] Hence rule explosion is eliminated for cases where the IRC rule table is additively separable. If an IRC rule table is not additively separable it is necessary to take an additional step to ensure that equality is attained. However, many IRC rule tables found in the literature are additively separable or are nearly additively separable in that only a minor number of IRC rules violate (3).

B. *Inseparable IRC Rule Tables*

When the IRC rule table is inseparable, additional steps are

[5] By rule, we refer to a statement of implication.

necessary to ensure that equality exists between IRC and URC systems. First, vector projections are selected that accurately represent a maximal number of the IRC rule table elements. For each element in the IRC rule table inaccurately represented by (3), it is necessary to add a *corrective term* to (11) that essentially relocates the corresponding consequent center of mass on the output axis.

The corrective term is expressed as

$$\theta(\overline{a}) = \frac{\left(\mathbf{F}(\overline{a}) - \sum_{i=1}^{P} v_i(a_i)\right)\prod_{j=1}^{P}\mu_{j,a_j}}{\sum_{a_1=1}^{N_1}\sum_{a_2=1}^{N_2}\cdots\sum_{a_P=1}^{N_P}\prod_{j=1}^{P}\mu_{j,a_j}}, \quad (15)$$

for a vector of indices $\overline{a}$, that specify the location of the consequent subset in the original P dimensional IRC rule table **F**. The corrective term undergoes steps similar to those required to progress from (5) to (11) to become

$$\theta(\overline{a}) = \frac{P\left(\mathbf{F}(\overline{a}) - \sum_{i=1}^{P} v_i(a_i)\right)\prod_{j=1}^{P}\frac{\mu_{j,a_j}}{\sum_{k=1}^{N_j}\mu_{j,k}}}{\sum_{i=1}^{P}\sum_{j=1}^{N_i}\frac{\mu_{i,j}}{\sum_{k=1}^{N_i}\mu_{i,k}}}. \quad (16)$$

Notice that this corrective term corresponds to a multi-antecedent rule where a product of normalized membership values yield a weight that gets applied at two locations on the output axis. The first term represents the correct location for the weight and the second term effectively removes the incorrectly positioned weight that is initially applied by (11).

Use of corrective terms allow for any arbitrary IRC rule table to be mapped to a URC system with the caveat that each corrective term adds one multi-antecedent rule. Therefore, it is important to consider the worst case, where a maximal number of corrective terms are required. The number of rules required in the worst case is given by

$$r_{URC} = (P-1) + \prod_{i=1}^{P} N_i . \quad (17)$$

Notice that the second term is the total number of rules in the original IRC rule table. The first term is the number of antecedents minus one. Thus, in the worst case, the URC system will actually contain more rules than the IRC system due to a conversion penalty of P-1 rules. In a future paper we address this issue and provide a method for obtaining a layered URC system that avoids rule explosion for inseparable rule tables [8].

Further, some have wondered if the URC system can incorporate an expert's perceived correlation between antecedents. It should be fairly obvious that this correlation does appear in the URC as a corrective term. Therefore, an expert's perceived correlation does appear in the URC system, but not without an increase in the computational burden. Finally, since any arbitrary IRC system can be converted to a URC system (which possibly contains corrective terms) and IRC fuzzy systems are proven universal approximators, URC fuzzy systems must be universal approximators as well [9-11].

III. Mapping an IRC to a URC: An Example

Consider a two antecedent, single consequent fuzzy system, where each antecedent has three subsets. Let the first antecedent be the quality of a student's Graduate Record Examination (GRE) with subsets of 'High', 'Medium', and 'Low'. Let the second antecedent be the quality of a student's undergraduate grade point average (GPA). In order to avoid confusion, different (albeit quirky) names are given to the subsets of GPA: 'Big', 'Average', and 'Small'. The consequent for this system shall represent the quality of an applicant to a graduate program and has subsets of 'Excellent', 'Very Good', 'Good', 'Fair', 'Poor', and 'Very Poor'. An IRC fuzzy inference engine that assigns a measure of goodness to graduate school applicants based on GPA and GRE scores is shown in Fig. 1. The consequent membership functions for this inference engine are shown in Fig 2. The antecedent membership functions are not shown, as they have no bearing on the mapping process provided that they are normalized as discussed previously.

Notice that the IRC rule table of Fig. 1 is additively separable. Therefore the IRC rule table can be directly transformed into a URC rule table via (12). Projection vectors are chosen to be the first row and column of the rule table minus an offset corresponding to one half of the element (1,1). As a result, the output axis is scaled by a factor of 2 and advanced by 10 units relative to the original consequent subsets (the advance is due to the choice of projection vectors). The rule table of the resulting URC is given in Fig. 3 and the consequent membership functions for the URC system are shown in Fig. 4.

		GPA		
		B	**A**	**S**
GRE	**H**	E	VG	G
	M	G	F	P
	L	F	P	VP

Fig. 1. The IRC rule table for the fuzzy inference engine

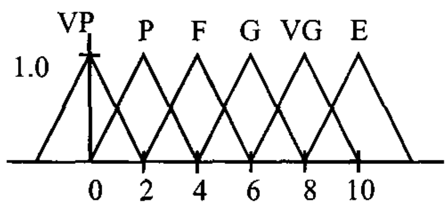

Fig. 2. Consequent membership functions for the IRC fuzzy inference engine.

GRE	H→E	M→G	L→F
GPA	B→E	A→VG	S→G

Fig. 3. The URC formulation of the IRC system given in Fig. 1. Notice that the number of rules has been reduced by 30%.

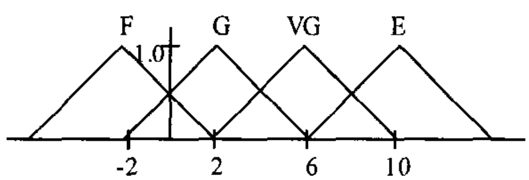

Fig. 4. The consequent membership functions for the URC fuzzy inference engine. Notice that the output axis has been subjected to an affine transformation.

The transformation from an IRC rule table to a URC rule table becomes slightly more involved when the IRC rule table is not additively separable. Consider the modified IRC rule table shown in Fig. 5. The IRC rules in this table are most accurately represented by the projection vectors from the previous example. However, the rule 'GRE is medium and GPA is average implies applicant quality is good,' does not obey (3). Therefore, it is necessary to add an extra multi-antecedent rule to the URC rule table. The additional rule removes the weight that is incorrectly applied to the consequent subset 'Fair' and instead applies the weight to the correct consequent subset 'Good'. However, as far as defuzzification is concerned, applying a negative weight at 'Fair' and a positive weight at 'Good' is the same as applying a single, positive weight at a new consequent subset 'X' which is found by subtracting the centers of mass of consequent subsets 'Good' and 'Fair'. Thus the URC system will contain a corrective rule and is shown in Fig. 6. The corresponding consequent membership functions are shown in Fig. 7.

		GPA		
		B	**A**	**S**
GRE	**H**	E	VG	G
	M	G	G	P
	L	F	P	VP

Fig. 5. An inseparable IRC rule table. The rule 'GRE is medium and GPA is average implies applicant quality is Good,' is not accurately represented by (3).

GRE	H → E	M → G	L → F
GPA	B → E	A → VG	S → G
Corrective Rule		M∩A → X	

Fig. 6. A URC rule table that implements the same fuzzy system as the inseparable IRC rule table given in Fig. 5. Note that an extra rule is required to handle the dependency between GRE and GPA.

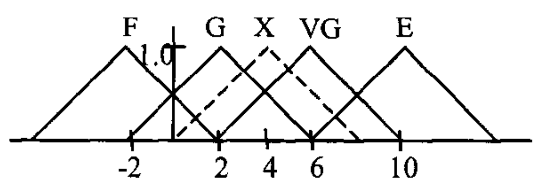

Fig. 7. The consequent membership functions for the URC rule table of Fig. 6. Notice the new consequent subset 'X' which is implied by the corrective rule.

IV. Conclusion

In conclusion, the mapping presented here is applicable for (potentially) non-square IRC rule tables with arbitrary antecedent membership functions, consequent membership functions with arbitrary (but identical) shape, and arbitrary defuzzification. Corrective terms are required for each rule in the IRC rule table that is not accurately represented by (2). As a matter of consequence, the URC must be a universal approximator due to the equality that exists between IRC and URC architectures.

Many fuzzy systems have additively separable IRC rule tables, and many more have nearly separable IRC rule tables. It is possible to dream up a fuzzy system with an IRC rule table that would require a large number of corrective terms.

One such example would be a table with XOR behavior. In this instance, the corresponding URC system actually requires more rules! In a future paper, we outline a method for circumventing rule explosion for inseparable rule tables [8].

If rule reduction is the primary goal, the IRC system should be examined carefully before it is converted to the URC architecture. For instance, one could design an IRC system where a maximum number of elements do not obey (3) do to a small difference in each instance. Thus, the analogous URC system actually contains more rules, yet a nearly identical URC system could perform the same task, with an exponential decrease in the number of rules if the consequent membership functions are placed along the output axis in a more efficient manner. Further, many IRC rule tables are inseparable only because designers intentionally limit the number of consequent subsets. In these cases, a single subset is repeated multiple times in the same row or column of the rule table (typically near the corners). Therefore, if an IRC rule table exhibits either of the two behaviors described above, it may be wise to modify the original IRC system before transforming it to a URC architecture or redesign the system directly as a URC fuzzy system.

Finally, we want to underscore the most important point of this paper. The mapping presented here is not as important as what the existence of the mapping means—URC systems can accomplish the same tasks as IRC systems, and most of the time URC systems can accomplish the tasks with fewer rules. Therefore we invite the reader to design a URC system and find out first hand how rule explosion is avoided.

ACKNOWLEDGMENT

We are thankful to Jerry M. Mendel for providing excellent feedback on a previous draft of this paper. His insightful comments helped us to clarify the presentation of this document.

REFERENCES

[1] 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.

[2] J. M. Mendel and Q. Liang, "Comments on 'Combinatorial rule explosion eliminated by a fuzzy rule configuration,'" *IEEE Trans. Fuzzy Sytems*, vol. 7, no. 3, pp. 396-373, June 1999.

[3] W.E. Combs, "Author's Reply," *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 3., June 1999.

[4] S. Dick and A. Kandel, "Comment on 'Combinatorial rule explosion eliminated by a fuzzy rule configuration,'" *IEEE Trans. Fuzzy Systems*, vol. 7, no. 4, pp. 475-477, Aug. 1999.

[5] W.E. Combs, "Author's Reply," *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 4., Aug 1999.

[6] J. M. Mendel and W. E. Combs, "The URC fuzzy system and universal approximation," submitted to *IEEE Trans. Fuzzy Systems* on 10/29/2002.

[7] B.-S. Chen, H.-J. Unang, and C.-S. Tseng, "Robust tracking enhancement of robot systems including motor dynamics: a fuzzy-based dynamic game approach," *IEEE Trans. Fuzzy Systems*, vol. 6, pp. 538-552, Nov. 1998.

[8] J. J. Weinschenk, R. J. Marks II, W. E. Combs, "Layered URC fuzzy systems: a novel link between fuzzy systems and neural networks," submitted to *Intl' Joint Conf. Neural Nets* on 1/29/2003.

[9] 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.

[10] L.-X. Wang, "Fuzzy systems are universal approximators," *Proc. IEEE Int'l. Conference on Fuzzy Systems*, San Diego, CA 1992, pp. 1163-1170.

[11] 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.