1000-9825/2002/13(10)1943-09 ©2002 Journal of Software 软 件 学 报 Vol.13, No.10

Hierarchical Mass-Assignment Fuzzy Systems of Two Types as Universal Approximation*

WANG Shi-tong[1,2], J.F. Baldwin[1], T.P. Martin[1]

[1](*Advanced Computing Research Center, Bristol University, UK*);

[2](*School of Information Technology, Southern Yangtze University, Wuxi* 214036, *China*)

E-mail: wxwangst@yahoo.com.cn

Received April 6, 1999; accepted October 8, 1999

Abstract: In this paper, hierarchical mass-assignment fuzzy systems of two types are presented, based on mass assignment theory. It is constructively proved that hierarchical mass-assignment fuzzy systems of these two types are also universal approximators. Because of the fact that the number of rules in type 1 hierarchical mass-assignment fuzzy systems increases linearly with the number of input variables and that fuzzy systems are added up to type 2 hierarchical mass-assignment fuzzy systems in terms of different accuracy requirements. These two types of systems can be effectively used to overcome rule-explosion problem, that is, the number of rules increases exponentially with the number of input variables

Key words: mass assignment theory; defuzzification; hierarchical fuzzy system; universal approximator

Mass assignment theory[1] is increasingly used in control, cased-based reasoning, data-browsing and search problems, meaning representation, deductive databases and function approximation, etc. Probability theory is adequate to deal with missing information and can be useful for generalization and simplification in certain cases but does not naturally help in the representation of vagueness of definition or for some essential forms of generalization. There are advantages of fuzzy set theory. Mass assignment theory combines the advantages of these two theories, avoiding their disadvantages.

The authors applied mass assignment theory and its implement language FRIL[1] to function approximation, obtaining very good approximation results. We have proved mass assignment approximation algorithm[2] is a universal approximator. However, for this mass assignment approximation algorithm, there also exists the 'curse of dimensionality', that is, the number of rules increases exponentially with the number of input variables. Hierarchical mass-assignment fuzzy systems presented here can effectively overcome this problem. In this paper, we will present two different hierarchical mass-assignment fuzzy systems and prove that they are universal approximators.

1 Essentials of mass assignment theory

In this section, we describe the basic ideas of the mass assignment theory. We will use a simple example to explain the ideas.

You are told that a weighted dice is thrown and the value is **small** where **small** is a fuzzy set defined as

small = 1/1 + 2/0.9 +3/0.4.

The prior probability for the dice is

1:0.1, 2:0.2, 3:0.3, 4:0.2, 5:0.1, 6:0.1

∗ **WANG Shi-tong** was born in 1964. He is a professor and doctorial supervisor. His research interests include AI, fuzzy systems, neural networks, pattern recognition and knowledge discovering.

Can we derive the distribution Pr(dice is i | dice is **small**) ?

What is Pr(dice is **about-2** | dice is **small**) where **about-2** is a fuzzy set defined as

about-2 = 1/0.3 +2/1 + 3/0.3.

The most fundamental question we must ask is what do we mean by **small**. What is the semantics of fuzzy sets? To answer this question we will use a voting model involving human voters. The world is not fuzzy. It is continuous and messy and we have to give labels to things we want to recognize as certain objects. We want to categorise and give labels to these categories. There will always be borderline cases. A particular object is neither a tree nor a bush but we do not have a label for it. We must therefore say that it is a borderline case but it may be more like a tree than a bush. We can therefore use graded membership in the nearest and most appropriate categories. We might say the object is a bush with a membership of 0.7 and a tree with a membership of 0.9. But what meaning can we give to this membership value?

Imagine that we have a representative set of people labelled 1 through 10. Each person is asked to accept or reject the dice score of x as **small.** They can believe x is a bordline case but they have to make a binary decision to accept or reject. We will take the membership of x in the fuzzy set **small** to be the proportion of persons who accept x as **small.** Thus we know that everyone accepted *1* as small. 90% of persons accepted 2 as small and 30% of persons accepted 3 as small. We only know the proportion of persons who accepted each score rather than the complete voting pattern of each person. We will assume that anyone who accepted x as being small will accept also any score lower than x as being small. With this assumption we can write down the voting pattern

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	persons
1	1	1	1	1	1	1	1	1	1	everyone accepts 1
2	2	2	2	2	2	2	2	2		90% accept 2
3	3									30% accept 3

Therefore 1 person accepts {1}, 6 persons accept {1, 2} and 3 persons accept {1,2,3}as being the possible sets of scores when they are the dice is small. If a member is drawn at random then the probability distribution for the set of scores this person will accept is

{1}:0.1, {1,2}:06, {1,2,3}:0.3

This is a probability distribution on the power set of dice scores and we will call this a mass assignment and write it as

$$m_{\text{small}} = \{1\}:0.1,\ \{1,2\}:0.6,\ \{1,2,3\}:0.3$$

We can determine the mass assignment very easily by using the method described in Ref.[1]. This mass assignment corresponds to a family of distributions on the set of dice scores. Each mass associated with a set of more than one element can be divided in some way amongst the elements of the set. This will lead to a distribution over the dice scores and there are an infinite number of ways in which this can be done.

Suppose we wish to give a unique distribution over the dice scores when we are told the dice value is small. How can we choose this distribution from the family of possible distributions arising from the mass assignment? To provide the least prejudiced distribution or the fairest distribution we would divide the mass amongst the elements of the set associated with them according to the prior for the dice scores. If this prior is unknown then we would use a local entropy concept and divide each mass equally among the elements of its set. The resulting distribution is called the least prejudiced distribution.

For the above case when we know the dice is **small** and has the prior given above we obtain the least prejudiced distribution

$$1:\ 0.1 + 1/3(0.6) + 1/6(0.3) = 0.35$$
$$2:\ 2/3(0.6) + 2/6(0.3) = 0.5$$

$$3: 3/6(0.3) = 0.15$$

Thus,

$$\Pr(\text{dice is } 1 \mid \text{dice is } \mathbf{small}) = 0.35$$
$$\Pr(\text{dice is } 2 \mid \text{dice is } \mathbf{small}) = 0.5$$
$$\Pr(\text{dice is } 3 \mid \text{dice is } \mathbf{small}) = 0.15$$

We will also use the notation

$$lpd_{\mathbf{small}} = 1{:}0.35,\ 2{:}0.5,\ 3{:}0.15$$

where *lpd* stands for the least prejudiced distribution.

This least prejudiced distribution plays a fundamental role in converting a probability distribution of a given feature to a fuzzy set. The FRIL language[2] can determine the least prejudiced distribution for any discrete or continuous fuzzy set. It can also determine the fuzzy set corresponding to any feature distribution treated as the least prejudiced distribution.

In the case of the prediction problem, the heads of the FRIL rules are of the form

$$(\text{value of } y \text{ is } B_i)$$

For a given case where the values of the features in the bodies of the rules are known, based on mass assignment theory, a solution will be inferred:

$$(\text{value of } y \text{ is } B)$$

where B is a fuzzy set on Y domain. We require a defuzzified value of y for our prediction. Firstly, the least prejudiced distribution lpd_B for the value of y is computed, and if g is a continuous fuzzy set, we use this distribution to determine the expected value of y. This expected value is taken as the defuzzified value. If g is a discrete fuzzy set then the defuzzified value is that value with the largest least prejudiced distribution probability. This method of defuzzification is justified by the voting model semantics.

Now, we consider another problem, that is, what is the probability of the dice value being **about-2** when we know it is **small** where **about-2** is a fuzzy set defined by

$$\mathbf{about\text{-}2} = 1/0.4 + 2/1 + 3/0.4$$

The mass assignment for the fuzzy set **about-2** is

$$m_{\text{about-2}} = \{2\}{:}0.6,\ \{1,2,3\}{:}0.4$$

We can use this mass assignment with the least prejudiced distribution for small to obtain a point value for Pr(dice value is **about-2** | dice value is **small**). From the least prejudiced distribution for **small** we obtain

$$\Pr(\{2\} \mid \mathbf{small}) = 0.5,\ \Pr(\{1,2,3\} \mid \mathbf{small}) = 0.35+0.5+0.15 = 1.$$

And we define the Pr(dice value is **about-2** | dice value is **small**) as

$$\begin{aligned}&\Pr(\text{dice value is } \mathbf{about\text{-}2} \mid \text{dice value is } \mathbf{small})\\ &= m_{\text{about-2}}(\{2\}) \Pr(\{2\} \mid \mathbf{small}) + m_{\text{about-2}}(\{1,2,3\}) \Pr(\{1,2,3\} \mid \mathbf{small})\\ &= 0.6 * 0.5 + 0.4 * 1 = 0.7.\end{aligned}$$

This process of determining Pr(**about-2** | **small**) is called point value semantic unification. There is also a interval semantic unification[2]. Both point and interval semantic unifications can be determined for both discrete and continuous fuzzy sets.

2 Mass-Assignment Fuzzy System

In this section, we will briefly describe mass-assignment fuzzy systems.

Definition 1. (1) By an $\&$-operation we mean a continuous function $f_{\&}$: $[0,1]\times[0,1]\to[0,1]$ that satisfies the following 4 properties:

- $f_{\&}(0, 0) = f_{\&}(0, 1) = f_{\&}(1, 0) = 0$, $f_{\&}(1, 1)=1$;

- $f_{\&}(a, b) = f_{\&}(b, a)$ for all a, b;
- $f_{\&}(a, b) \le a$ for all a, b;
- if $a > 0$ and $b > 0$, then $f_{\&}(a, b) > 0$.

(2) By an $\vee$-operation, we mean a continuous function $f_{\vee}$: $[0, 1]\times[0, 1] \to [0, 1]$ that satisfies the following 3 properties:

- $f_{\vee}(0, 0) = 0$, $f_{\vee}(0, 1)= f_{\vee}(1, 0) = f_{\vee}(1, 1) =1$;
- $f_{\vee}(a, b) = f_{\vee}(b, a)$ for all a, b;
- $f_{\vee}(a, b) \ge a$ for all a, b.

(3) By a defuzzification procedure F, we mean a mapping that transforms a membership function $\mu(x)$ into a number and satisfies the following properties:

- if $\mu(x) = 0$ for all $x \in (-\infty, a)$ then $F(\mu) \ge a$;
- if $\mu(x) = 0$ for all $x \in (-\infty, a)$ then $F(\mu) > a$;
- if $\mu(x) = 0$ for all $x \in (a, \infty)$ then $F(\mu) \le a$;
- if $\mu(x) = 0$ for all $x \in [a, \infty]$ then $F(\mu) < a$.

Theorem 1. $f_{\&}$ and $f_{\vee}$ are the extensions of T-norm and S-norm respectively.

The correctness of this theorem is obvious.

In a mass-assignment fuzzy system, the fuzzy rule is given using FRIL:

$$(y \text{ is } B_i\ ) \text{ iff } (x_1 \text{ is } A_i^1) \text{ and ... and } (x_n \text{ is } A_i^n\ ):(1,1)(0,0)$$

where A_i^r is a fuzzy set defined on X_r, its membership function is $A_i^r(x_r)$, B_i is a fuzzy set defined on Y, its membership function is $B_i(y)$, i =1,2,...,M, $B_1(y)+B_2(y)+\ldots+B_M(y)=1$, and M represents the number of fuzzy logic rules in the fuzzy knowledge base.

Given the values of input variables x^*_1, x^*_2,...,x^*_n, for ith rule, we determine the conditional probabilities $Pr(X \text{ is } A_i^r \mid X \text{ is } x^*_i)$ by using point value semantic unification in mass assignment theory, where i=1,2,...,n, r=1, 2,...,M.

$$Pr(X \text{ is } A_i^r \mid X \text{ is } x^*_i) = A_i^r(x_r\ )$$

We define

$$P_i = A_i^1(x_1)\quad f_{\&}\quad A_i^2(x_2\ )\quad f_{\&} \ldots f_{\&}\quad A_i^r(x_r)\quad f_{\&}\quad B_i(y)$$

Then we use $f_{\vee}$ to combine the above M fuzzy logic rules, thus, we obtain the fuzzy set B and its membership function $B(y)$ of output variable y :

$$B(y) = P_1\, f_{\vee}\, P_2\, f_{\vee} \ldots f_{\vee}\, P_M$$

As a result, the real output of the above mass-assignment fuzzy system y is defined as

$$y=\int\ y\ lpd_B\ (y)\ \mathrm{d}y$$

where $lpd_B(y)$ represents the least prejudiced distribution of $B(y)$ in mass assignment theory.

Theorem 2[2]. When $A_i^r(x_r)$ (r=1,2,...,n, i=1,2,...,M) satisfy the following: (1) continuous; (2) > 0 in some interval (a b); (3) = 0 outside the interval (a, b), then the above mass-assignment fuzzy system is a universal approximator.

3 Universal Approximation by Type 1 Hierarchical Mass-Assignment Fuzzy Systems

In this section, we first define type 1 hierarchical mass-assignment fuzzy systems, then investigate the property of their universal approximation.

Type1 mass-assignment fuzzy system is shown in Fig.1. In Fig.1, we use m-fuzzy system to represent a mass-assignment fuzzy system for simplicity. We see that this n-input hierarchical mass-assignment fuzzy system

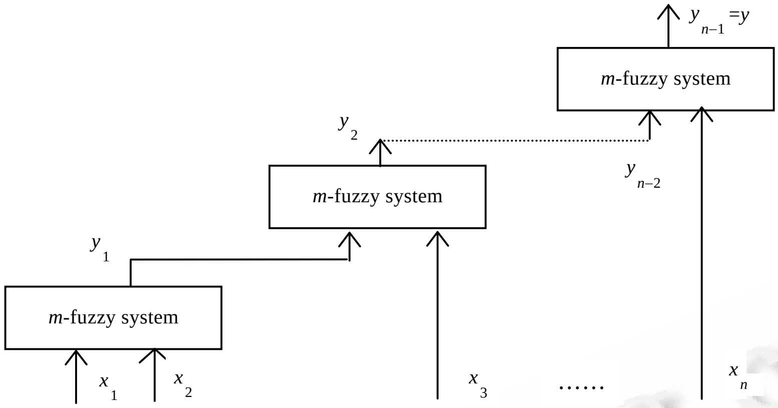

Fig.1 Type 1 hierarchical fuzzy mass-assignment system

comprises $n-1$ low-dimensional m-fuzzy systems, with each low-dimensional m-fuzzy system having two inputs. If we define L fuzzy sets for each variable, including the internal variables $y_1,y_2,\ldots,y_{n-2}$, the total number of rules is $(n-1)L^2$ which is a linear function of the number of input variables n, thus we have

Theorem 3. The number of rules in type 1 hierarchical mass-assignment fuzzy system increases linearly with the number of input variables.

In the above hierarchical mass-assignment fuzzy system, for the first level, fuzzy logic rules are denoted as:

$$\text{value of } y_1 \text{ is } B_{1j} \text{ IFF } x_1 \text{ is } A_{1j}^{\ 1} \text{ and } x_2 \text{ is } A_{1j}^{\ 2}\text{:}(1, 1)(0, 0)$$

where $B_{11}(y_1)+B_{12}(y_1)+\ldots+B_{1m1}(y_1)=1$ ($m1$ denotes the number of rules at the first level, mi denotes the number of rules at level i). For the ith level, fuzzy logic rules are denoted as:

$$\text{value of } y_i \text{ is } B_{ij} \text{ IFF } x_{i+1} \text{ is } A_{ij}^{\ i+1} \text{ and } y_{i-1} \text{ is } C_{i-1j}\text{:}(1, 1)(0, 0)$$

where $B_{i1}(y_i)+B_{i2}(y_i)+\ldots+B_{imi}(y_i)=1$, when $i = n-1$, $y_i=y$.

Now, we use the constructive proof method to prove that the above hierarchical mass-assignment fuzzy system is a universal approximator.

Theorem 4. Type 1 hierachical mass-assignment fuzzy system is a universal approximator.

Proof. Suppose $U = X_1\times X_2\times\ldots\times X_n$ is compact, then there exists a finite $(\delta/2)$-net, i.e., a finite set of points $\boldsymbol{x}^1$, $\boldsymbol{x}^2,\ldots,\boldsymbol{x}^K\in U$, such that for any $\boldsymbol{x}\in U$, there exists a j for which $\rho(\boldsymbol{x}, \boldsymbol{x}^j)\le\delta/2$. Let us fix such a net.

Suppose $g_1(x_1,x_2)$ belongs to $D_1(X_1\times X_2)$, $g_1(x_1,x_2,x_3)$ belongs to $D_1(X_1\times X_2\times X_3),\ldots,g(\boldsymbol{x})$ belongs to $D(U)$.

(1) Firstly, suppose $\varepsilon > 0$, for the first level, let us prove that there exists an m-fuzzy system such that if $\rho(\boldsymbol{x},\boldsymbol{x}^j)\le\delta/2$ then $y_1\in[g_1(x_1,x_2)-\varepsilon, g_1(x_1,x_2)+\varepsilon]$. Let us construct such an m-fuzzy system.

At the first level, each rule will take the following form (for point $\boldsymbol{x}_j$):

$$\text{value of } y_1 \text{ is } B_{1j} \text{ IFF } x_1 \text{ is } A_{1j}^{\ 1}(x_1) \text{ and } x_2 \text{ is } A_{1j}^{\ 2}(x_2)\text{:}(1,1)(0,0)$$

and
$$\sum_{j=1}^{m1} B_{1j}(y_1) = 1,$$

where the corresponding membership functions are defined as follows. $A_{1j}^{\ i}(x)=\mu_x((x-x_i^{\ j})/\delta)$, $B_{1j}(y)=\mu_y((y-y_1)/(\varepsilon/2))$, where $y_1=g_1(x_1,x_2)$ and $\mu_x(x)=\mu_x(x(b-a)/2+(a+b)/2)$ is a function that is >0 only for $x\in[-1,1]$ while μ_y is defined similarly .

Let us denote

$$\mu_{D1}(y_1)=f_\vee^{\ 1}(p_1^{\ 1},\ldots, p_j^{\ 1},\ldots, p_{m1}^{\ 1})$$
$$p_j^{\ 1} = f_\&^{\ 1}(A_{1j}^{\ 1}(x_1), A_{1j}^{\ 2}(x_2), B_{1j}(y_1)) \tag{1}$$

where $f_\vee^{\ 1}$ and $f_\&^{\ 1}$ are $f_\vee$ and $f_\&$ operations respectively.

Now, let us show that $\mu_{D1}(y_1)$ is not identically 0. Since we chose the set $\{\boldsymbol{x}^j\}$ as a $\delta/2$-net, there exists a j such that $|x_i-x_i^j|\le\delta/2$ for i=1,2, therefore $|x_i-x_i^j|\le\delta$ which means that $x_i\in(x_i^j-\delta, x_i^j+\delta)$ and so $A_{1j}^1(x_1)$ and $A_{1j}^2(x_2)>0$. If we take $y_1=g_1(x_1^j, x_2^j)$, we conclude that $B_{1j}(y_1)>0$. Since we required that $f_\&^1(a,b)>0$ if $a>0$ and $b>0$, we can conclude that

$$p_j^1 = f_\&^1(A_{1j}^1(x_1),\quad A_{1j}^2(x_2), B_{1j}(y_1)) > 0$$

Since $f_\vee^1(a,b)>\max(a,b)$, we conclude that for this y_1

$$\mu_{D1}(y_1)= f_\vee^1(p_1^1,..., p_j^1,..., p_{m1}^1) \ge p_j^1 > 0$$

So $\mu_{D1}(y_1)$ is not identically 0.

We now prove that if $|y_1-g_1(x_1,x_2)|>\varepsilon$, then $\mu_{D1}(y_1)= 0$ by showing that in this case $p_1^1=...=p_j^1=...=p_{m1}^1=0$, so

$$\begin{aligned}\mu_{D1}(y_1)&= f_\vee^1(p_1^1, ... , p_j^1,..., p_{m1}^1)\\ &= f_\vee^1(0, ... , 0, ..., 0)\\ &= 0.\end{aligned}$$

Now, let us take an arbitrary j from the interval 1 to $m1$ and prove that $p_j^1=0$. Indeed, since

$$p_j^1=f_\&^1(A_{1j}^1(x_1), A_{1j}^2(x_2), B_{1j}(y_1)) > 0$$

$f_\&^1(0,p)=0$, then only possibility for p_j^1 to be positive is $A_{1j}^1(x_1)$, $A_{1j}^2(x_2)$, $B_{1j}(y_1) > 0$ respectively. In terms of $A_{1j}^1(x_1)$, $A_{1j}^2(x_2)$ is positive only for $|x_i- x_i^j|\le\delta$. In this case, by the choice of δ, $|g_1(x_1,x_2)-g_1(x_1^j, x_2^j)|=|y_1-g_1(x_1, x_2)|\le\varepsilon/2$, but we assumed that $|y_1-g_1(x_1^j,x_2^j)|>\varepsilon$, therefore $|y_1-y_1^j|\ge|y_1-g_1(x_1,x_2)|-|y_1^j-g_1(x_1,x_2)|>\varepsilon/2$ and $B_{1j}(y_1)=0$.

For every (x_1,x_2), therefore, either one of terms $A_{1j}^1(x_1)$ and $A_{1j}^2(x_2)$ is 0 or they are all positive, in which case $B_{1j}(y_1)=0$. In both cases, $p_j^1=0$, so $f_\vee^1(p_1^1,..., p_j^1,...,p_{m1}^1)=f_\vee^1(0,...,0,...,0)=0$ for all y_1 outside an interval $[g_1(x_1,x_2)-\varepsilon, g_1(x_1,x_2)+\varepsilon]$ and $y_1=F_1(\mu_{D1})$ belongs to this interval, i.e., there exists an m-fuzzy system at the first level such that $y_1\in[g_1(x_1,x_2)-\varepsilon,g_1(x_1,x_2)+\varepsilon]$ if $\rho(\mathbf{x},\mathbf{x}^j)\le\delta/2$, where F_1 is a least prejudice distribution *lpd* of μ_{D1}, which is a defuzzification procedure.

(2) Secondly, let us construct an m-fuzzy system at the second level such that $y_2\in[g_2(x_1,...,x_3)-\varepsilon, g_2(x_1,...,x_3)+\varepsilon]$ if $\rho(\mathbf{x}, \mathbf{x}^j)\le\delta/2$.

At the second level, each rule can be represented in the following form:

value of y_2 is B_{2j} IFF x_3 is $A_{2j}^3(x_3)$ and y_1 is $C_{1j}(y_1)$,

and
$$\sum_{j=1}^{m1} B_{2j}(y_1) = 1.$$

In this case,

$$p_j^2 = f_\&^2(A_{2j}^3(x_3), C_{1j}(y_1), B_{2j}(y_1)). \tag{2}$$

We define $A_{2j}^3(x)= \mu_{x1}((x-x_3^j)/\delta)$, $C_{1j}(y_1)= \mu_{y1}((y_1-g_1(x_1,x_2))/\varepsilon)$, $B_{2j}(y)=\mu_{y2}((y-y_2)/(\varepsilon/2))$, where $y_2=g_2(x_1,...,x_3)$ and $\mu_{x1}(x)=\mu_x(x(d-c)/2+(c+d)/2)$ is a function that is > 0 only for $x\in[-1,1]$ while μ_{y1} and μ_{y2} are defined similarly.

Like the above proof in step (1), we can easily show the above defined m-fuzzy system can satisfy: $y_2\in[g_2(x_1,...,x_3)-\varepsilon, g_2(x_1,...,x_3)+\varepsilon]$ if $\rho(\mathbf{x}, \mathbf{x}^j)\le\delta/2$. By comparing (1) and (2), in terms of the above proof in step (1), we only need to prove that $C_{1j}(y1) > 0$ if $\rho(\mathbf{x}, \mathbf{x}^j)\le\delta/2$.

In the proof of step (1), we conclude $y_1\in[g_1(x_1,x_2)-\varepsilon, g_1(x_1,x_2)+\varepsilon]$ if $\rho(\mathbf{x},\mathbf{x}^j)\le\delta/2$, i.e., if $\rho(\mathbf{x}, \mathbf{x}^j)\le\delta/2$, $|y_1-g_1(x_1,x_2)|\le\varepsilon$, i.e.,

$$0<=|y_1-g_1(x_1, x_2)|/\varepsilon\le 1.$$

Thus, we have $C_{1j}(y_1) =\mu_{y1}((y_1-g_1(x_1,x_2))/\varepsilon) > 0$.

(3) Similarly, for other levels, the above conclusions hold. Now let us consider the last level. At this level, we construct an m-fuzzy system such that $y_{n-1}=y\in[g(\mathbf{x})-\varepsilon, g(\mathbf{x})+\varepsilon]$, if $\rho(\mathbf{x}, \mathbf{x}^j)\le\delta/2$. In this m-fuzzy system, each rule will take the following form:

value of y is B_{n-1j} IFF x_n is $A_{n-1j}^n(x_n)$ and y_{n-2} is $C_{n-2j}(y_{n-2})$:(1,1)(0,0) and $$\sum_{j=1}^{m1} B_{n-1j}(y) = 1$$

we define $A_{n-1j}{}^{n}(x)=\mu_{xn-1}((x-x_n{}^{j})/\delta)$, $C_{n-2j}(y_{n-2})=\mu_{yn-2}\ ((y_{n-2}-g_1(x_1,...,x_{n-1}))/\varepsilon)$, $B_{n-1j}(y)=\mu_{yn-1}((y-y_{n-1})/(\varepsilon/2))$, where $y_{n-1}=g_{n-1}(x_1,...,x_n)$ and $\mu_{xn-1}(x)=\mu_x(x(e-f)/2+(e+f)/2)$ is a function that is >0 only for $x\in[-1,1]$ while μ_{yn-1} and μ_{yn-2} are defined similarly.

Also, we can easily show that the above m-fuzzy system at the last level can satisfy $y\in[g(\boldsymbol{x})-\varepsilon, g(\boldsymbol{x})+\varepsilon]$, if $\rho(\boldsymbol{x}, \boldsymbol{x}^j)\le\delta/2$. Thus, this theorem is completely proved.

It should be pointed out that for general hierarchical mass-assignment fuzzy system, i.e., leve i has n_i input variables instead of 2 (level 1) or 1 (other levels) input variables, the above proof can easily be extended to this case, the above theorem also holds.

4 Universal Approximation by Type 2 Hierarchical Mass-Assignment Fuzzy Systems

In Ref.[6], Yager presented another hierarchical type fuzzy system called HPS (Hierarchical Prioritized Structure). Based on the idea of HPS, we define type 2 hierarchical mass-assignment fuzzy system, in which we allow a hierarchical representation of the rules along with a new aggregation technique enabling us to aggregate the information provided at different levels of hierarchy. This newly-defined system enables us to introduce exceptions to more general rules by giving them a priority and introducing them at a higher level in the hierarchy. When the solution of level n does not satisfy the accuracy requirement, we add level $n-1$ to this system. In this way we construct type 2 hierarchical mass-assignment fuzzy system until the accuracy requirement is reached. Thus, the rules are organized in the hierarchy so that the number of rules is determined according to practical requirements.

Now, let us describe type 2 hierarchical mass-assignment fuzzy system. Figure 2 gives its structure. Assume we have a system we are modeling with input vector $\boldsymbol{x}=(x_1,x_2,\ldots,x_n)$. At each level of this type 2 system, we have a collection of fuzzy rules in FRIL. Thus, for level i, we have a collection of mi rules:

value of y is B_{ij} IFF x_1 is $A_{ij}{}^1(x_1)$ and ... and x_n is $A_{ij}{}^n(x_n)$: (1, 1)(0, 0).

According to the same method as that in Section 2, for level i, we have

$$B_i\,(y)=f_{\vee}(p_{i1},\ldots, p_{i2},\ldots, p_{imi})$$

where $p_{ij}=f_{\&}(A_{ij}{}^1(x_1), ..., A_{ij}{}^n(x_n),B_{ij}(y))$.

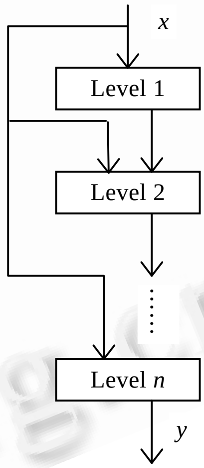

Fig.2 Type 2 hierarchical mass-assignment fuzzy system

In type 2 hierarchical mass-assignment fuzzy system, the output of the ith level, G_i, is obtained by combining the output of the previous level G_{i-1} with $B_i(y)$, by using Hierarchical Updation aggregation operator HU. The output of the last level G_n is then considered as the fuzzy output of this type 2 system. In addition, initialize the process by assigning $G_i=\varnothing$.

The HU aggregation operator is defined as

$$G_i(y)=G_{i-1}(y)+(1-\alpha_{i-1})B_i(y) \tag{3}$$

where $\alpha_{i-1}=\text{Max}\,[G_{i-1}(y)]$, i.e., the largest membership grade in G_{i-1}. When $i=n$, we use $lpd(G_n(y))$ as the real output of type 2 hierarchical mass-assignment fuzzy system.

We can use triangular norms S and T to expand formula (3), and thus we get

$$G_i(y)=S(G_{i-1}(y), T((1-\alpha_{i-1}),B_i(y))) \tag{4}$$

Furthermore, we use $f_{\vee}$ and $f_{\&}$ in the above formula (4) instead of S, T to get GHU (Generalized HU) aggregation operator, and we have

$$G_i(y)=f_{\vee}(G_{i-1}(y), f_{\&}((1-\alpha_{i-1}), B_i(y))). \tag{5}$$

Now, we will prove that type2 hierarchical mass-assignment fuzzy system is also universal approximator.

Theorem 5. Type 2 hierarchical mass-assignment fuzzy system is a universal approximator.

Proof. We use inductive method to prove this theorem.

Suppose $U= X_1\times X_2\times \ldots\times X_n$ is compact, $g(\boldsymbol{x})$ belongs to $D(U)$.

(1) i=1, $G_1(y)=B_1(y)$, thus, type 2 hierarchical mass-assignment fuzzy system becomes a mass-assignment fuzzy system. We can easily prove this assertion.

(2) Suppose when $i=i-1$, this theorem holds. Now we need to prove that this theorem also holds when $i= i$.

Because of the hypothesis that this theorem holds for $i=i-1$, in terms of proof step (1) in theorem 4, there exists a finite $(\delta/2)$-net, i.e., a finite set of points $\boldsymbol{x}^1, \boldsymbol{x}^2,\ldots,\boldsymbol{x}^K\in U$, such that for any $\boldsymbol{x}\in U$, there exists a j for which $\rho(\boldsymbol{x}, \boldsymbol{x}^j)\le\delta/2$, and $\varepsilon>0$ and $G_{i-1}(y)=0$ for all points outside the interval $[g(\boldsymbol{x})-\varepsilon, g(\boldsymbol{x})+\varepsilon]$, and $lpd(G_{i-1}(y))\in[g(\boldsymbol{x})-\varepsilon, g(\boldsymbol{x})+\varepsilon]$, if $\rho(\boldsymbol{x}, \boldsymbol{x}^j) \le\delta/2$.

Similarly, for this $(\delta/2)$-net, we can construct a mass-assignment fuzzy system at level i, such that for any $\boldsymbol{x}\in U$, there exists a j for which $\rho(\boldsymbol{x}, \boldsymbol{x}^j)\le\delta/2$ and $\varepsilon>0$ and $B_i(y)$ =0 for all points outside the interval $[g(\boldsymbol{x})-\varepsilon, g(\boldsymbol{x})+\varepsilon]$, and $lpd(B_i(y))\in[g(\boldsymbol{x})-\varepsilon, g(\boldsymbol{x})+\varepsilon]$, if $\rho(\boldsymbol{x}, \boldsymbol{x}^j) \le\delta/2$.

Now, let us consider the following case. For all outside an interval $[g(\boldsymbol{x})-\varepsilon, g(\boldsymbol{x})+\varepsilon]$, when $\rho(\boldsymbol{x}, \boldsymbol{x}^j)\le\delta/2$, $B_i(y)=0$, $G_{i-1}(y) = 0$. In terms of formula (5), we have

$$\begin{aligned} G_i(y) &= f_{\vee}(0, f_{\&}((1-\alpha_{i-1}), 0)) \\ &= f_{\vee}(0, f_{\&}(0, 0)) \qquad (\alpha_{i-1}=0 \text{ at this time}) \\ &=f_{\vee}(0, 0) = 0. \end{aligned}$$

Because $G_i(y) = f_{\vee}(G_{i-1}(y), f_{\&}((1-\alpha_{i-1}), B_i(y)))$ is a membership function about y in the interval $[g(\boldsymbol{x})-\varepsilon,g(\boldsymbol{x})+\varepsilon]$, hence, $lpd(B_i(y))\in[g(\boldsymbol{x})-\varepsilon, g(\boldsymbol{x})+\varepsilon]$, if $\rho(\boldsymbol{x}, \boldsymbol{x}^j)\le\delta/2$.

Thus, in terms of inductive method, this theorem is completely proved.

For hierarchical prioritized system HPS presented by Yager, we have:

Theorem 6. HPS is also universal approximator.

Proof. Because formula (4) is a special case of formula (5), *lpd* and GOA are defuzzification methods, so HPS is a special case of type 2 hierarchical mass-assignment fuzzy system. Therefore, this theorem holds.

It should be pointed out that we can use the method in Ref.[6] to construct a type 2 hierarchical mass-assignment fuzzy system from rules.

5 Conclusions

If the number of input variables increases, due to suffering from the curse of dimensionality, a fuzzy system will become increasingly intractable. In this paper, based on mass assignment theory, we present hierarchical mass-assignment fuzzy systems of two types. Theoretical research results show that they are also universal approximator, which are important for their practical applications.

Finally, we should point out, many scholars, such as Dr. O. Huwendiek, Dr. Wang Lixing, Dr. R.R. Yager, have investigated hierarchical fuzzy systems in recent years. R.R. Yager presented HPS system, however, he did not prove that it is a universal approximator.

All hierarchical fuzzy systems presented by them are the special cases of type 1 and type 2 systems here. Hence, type 1 and type 2 hierarchical mass-assignment fuzzy systems are the most generalized hierarchical fuzzy systems so far. Further research work is to investigate the learning techniques of them.

Acknowledgement This work is supported by British Royal Society.

References:

[1] Baldwin, J.F., *et al*. FRIL-Fuzzy and Evidential Reasoning in AI. Research Studies Press, 1995.

[2] Baldwin, J.F., *et al*. Knowledge from Data Using Fril and Fuzzy Methods. London: John Wiley and Sons Press, 1996.

[3] Wang, Shi-tong. Neural Fuzzy Systems and Their Applications. Beijing: Publishing House of Beijing University Aeronautics and Astronautics, 1998 (in Chinese).

[4] Wang, Shi-tong. Fuzzy Systems and Their Programming Technologies. Shanghai: Press of Shanghai Science and Technologies, 1998 (in Chinese).

[5] Wang, L.X., *et al*. Fuzzy system, fuzzy basis function and OLS. IEEE Transactions on Neural Networks, 1992,3(5):1173~1189.

[6] Yager, R.R. Constructing prioritized fuzzy models. In: Proceedings of the International Workshops on AI in IJCAI'95. Kluwier Publisher, 1995. 1~14.

[7] Huwendiek, O., *et al*. Function approximation with decomposed fuzzy systems. Journal of Fuzzy Sets and Systems, 1999,(101): 273~286.

[8] Wang, L.X. Universal approximation by hierarchical fuzzy systems. Journal of Fuzzy Sets and Systems, 1998,(93):223~230.

附中文参考文献**:**

[3] 王士同.神经模糊系统及其应用.北京航空航天大学出版社, 1998.

[4] 王士同.模糊系统及其编程技术.上海:上海科技出版社,1998.

两类层次模糊 **Mass-Assignment** 系统是全局逼近器

王士同[1,2], J.F. Baldwin[1], T.P. Martin[1]

[1](英国 Bristol 大学 先进计算研究中心,英国);

[2](江南大学 信息学院,江苏 无锡 214036)

摘要: 基于 mass assignment 理论,提出了两类层次模糊 mass assignment 系统,并运用构造性证明过程证明了其全局逼近性质.由于类型 1 层次模糊系统的规则数与输入变量数呈线性关系,类型 2 层次模糊系统按逼近精度要求引入子模糊系统,因此,此两类层次模糊 mass assignment 系统可被用来有效地克服模糊规则爆炸问题,即所谓的规则数与输入变量数呈指数关系问题.

关键词: mass assignment 理论;去模糊化;层次模糊系统;全局逼迫器

中图法分类号: TP18 文献标识码: A