Fuzzy Systems As Universal Approximators

V. Suresh Kumar
Principal, Kottayam Institute of Technology & Science, Kottayam, India.
lakshmikumarvs@yahoo.co.in

Abstract: **-Neural systems can approximate the unknown fuzzy values from training data. This has been studied and established by researchers K. Hornik et al., B. Kosko, L. Wang and J. M. Mendel. In this research paper, discussions about fuzzy systems as universal approximators have been done to a greater intensity. Functions with fuzzy patches have been considered for approximation. The approximation depends on the size of the fuzzy patches.**

Index terms: ***Approximation, fuzzy rules, Stone-Weierstrass theorem, additive fuzzy systems, defuzzification, extension principle***

I. INTRODUCTION

An additive fuzzy system can uniformly approximate any real continuous function on a compact domain to any degree of accuracy. An additive fuzzy system approximates the function by covering its graph with fuzzy patches in the input-output state space and averaging patches that overlap. Each fuzzy rule defines a fuzzy patch and connects commonsense knowledge with state-space geometry. Neural systems can approximate the unknown fuzzy patches from training data. At the local level the neural system approximates and tunes the fuzzy rules. At the global level the rules or patches approximate the function.

II. FUZZY APPROXIMATION AS A FUZZY COVERING

A fuzzy system approximates a function by covering its graph with fuzzy patches and averaging patches that overlap. The approximation improves as the fuzzy patches grow in number and shrink in size.

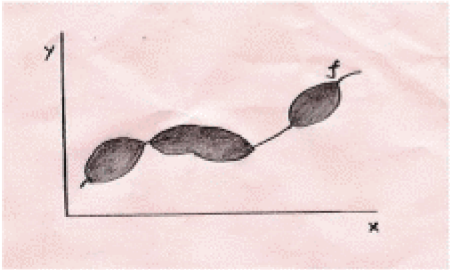

Fig. 1(a)

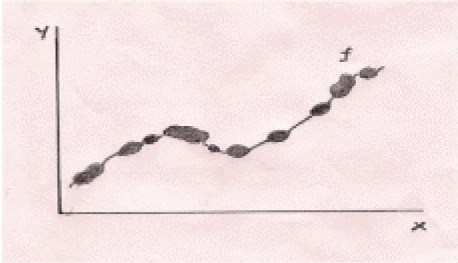

Fig. 1(b)

Fig. 1 *shows how fuzzy patches in the input-output product space* $X \times Y$ *cover the real function* $f : X \to Y$.

In Fig. 1(a), a few large patches approximate f. In Fig. 1(b), several smaller patches approximate f in a better way. On one hand, the approximation improves as we add more and more small patches. On the other hand, the storage cost and complexity increase. This brief contribution gives the algebraic details of the fuzzy approximation.

A fuzzy system is a set of if-then fuzzy rules that maps inputs to outputs. Each fuzzy rule defines a fuzzy patch in the input-output state space of the function.

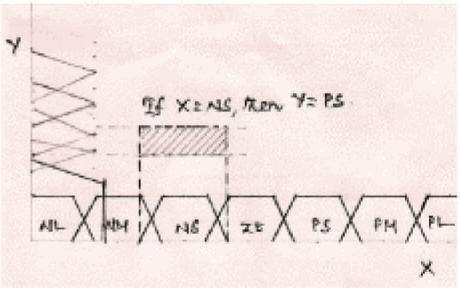

Fig.2

Fig. 2 shows the fuzzy rule "if X is Negative Small, then Y is Positive Small" as the Cartesian product NS × PS of "fuzzy" [2] or multi-valued sets NS and PS.

A 3-D plot would show the fuzzy patch NS × PS as a barn-like structure that rises up from its rectangular base. Each input belongs to some degree of each input fuzzy set. So each input fires all the fuzzy rules of some degree. Experts state the fuzzy rules or a neural or statistical system learns them from sample data. Experts and algorithms can give different sets of fuzzy rules and so give different approximations of the function.

Next, we show that a fuzzy system can approximate any continuous real function defined on a compact (closed and bounded in R^n) domain and show that even a bivalent expert system can uniformly approximate a bounded measurable function.

The fuzzy systems have a feedforward architecture that resembles the feedforward multilayer neural systems used to approximate functions [4]. The uniform approximation of continuous functions allows us to replace each continuous fuzzy set with a finite discretization or a point in a unit hypercube [7] or "fuzzy cube" of higher dimension.

Hornik and White [4] and others have used the Stone-Weierstrass theorem of functional analysis [9] to show uniform convergence of neural networks.

Radial basis nets can sum up Gaussian functions and also uniformly approximate continuous functions on compact sets [3]. Additive fuzzy systems [7] with Gaussian fuzzy sets [11] define radial basis nets and so also act as uniform approximators. The theorem below also proves so directly since it holds for all additive systems. The constructive proof below shows how to use neural systems to learn rules and how to let the rules or patches change with time to track a non-stationary function.

III. ADDITIVE FUZZY SYSTEMS

Inputs fire the if-part A_j of all fuzzy rules "if $X = A_j$, then $Y = B_j$," and give scaled sum B'_j as output. Earlier fuzzy systems combined the output fuzzy sets B'_j with pair-wise maximum in accord with the so-called "extension principle'' [2].

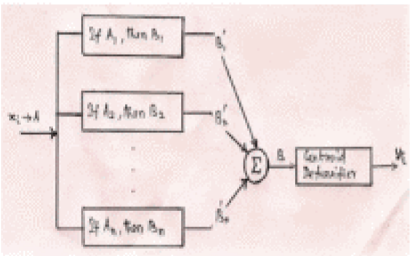

Fig. 3
Additive fuzzy systems sum the outputs

$$B = \sum_{j=1}^{m} \omega_i B'_j \qquad (1)$$

where B'_j is a sigmoidal signal that may be generalized sigmoidal, left sigmoidal or right sigmoidal. A typical situation is illustrated in fig. 3.

Combining the scaled or "fired" consequent fuzzy sets B'_1, …….. B'_m In Fig. 3 with pairwise maximum gives the envelope of the fuzzy sets and tends to the uniform distribution [5].

Max combination ignores overlap in the fuzzy sets B'_j. Sum combination adds overlap to the peak of B. When the input changes slightly, the additive output B changes slightly. The max- combined output may ignore small input changes since for large sets of rules most change occurs in the overlap regions of the fuzzy sets B'_j . Here the overlap problem arises since the centroid tends to stay in the same state for small changes in input.

Centroid defuzzification compounds the max problem. An integrable multivalued set function m_B : Y → [0,1] gives the centroidal output y_i or F(x) as

$$y_i = \frac{\int_{-\infty}^{\infty} y m_B(y) dy}{\int_{-\infty}^{\infty} m_B(y) dy} \qquad (2)$$

We can replace the integral in (2) with small discrete sums indexed only by a number of fuzzy sets that quantize the fuzzy variables [6]. This eliminates both the need to approximate the centroid and its computational burden. Digital VLSI hardware can then implement (2) or an analog VLSI chip can implement it with a simple follower-aggregator circuit [8]. In the limit as the number of combined output sets B'_j grows, centroidal defuzzification may tend to coincide with mode defuzzification since the centroid and mode coincide for a symmetric unimodal distribution.

IV. SIGMOIDAL FUZZY FUNCTION APPROXIMATION

Some inference reduces to the activation function that is a function of a weighted sum

$$B = \phi\left(\sum_{j=1}^{m} w_j a_i^j B_j\right)$$

(3)

where a_i^j is the degree to which input x_i, belongs to fuzzy set A_j, in the rule or patch $A_j \times B_j$. Additive fuzzy systems can approximate a function $f: X \to Y$ by shrinking these patches in size and increasing them in number [7].

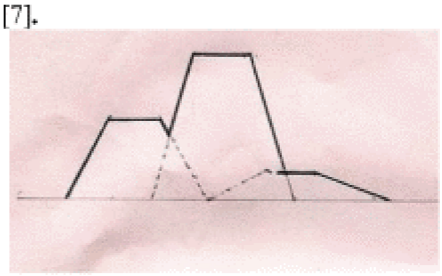

Fig. 4 (a)

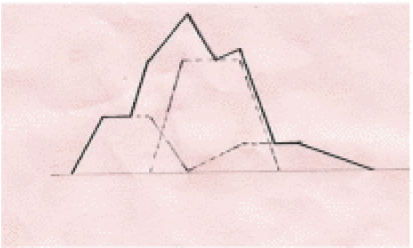

Fig. 4 (b)

Equation (3) gives a simple approximation theorem. Suppose $f: X \to Y$ is measurable and bounded. Then we can view B as a function of a simple function if the fuzzy sets B_j are non-fuzzy sets. We replace triangles, trapezoids, and other fuzzy sets with rectangles. Simple functions s: $X \to Y$ map X into finitely many values of Y. A simple function equals a finite sum of weighted indicator functions. Choose the fit values a_i^j as the weights and let the non-fuzzy sets or rectangles partition X.

It is not true that an sigmoidal fuzzy system with multivalued sets converges uniformly to f. So in practice the result gives no error bound. The problem arises because triangles, trapezoids, and other fuzzy sets need not converge uniformly to rectangles. It can be easily understood thro' fig. 4(a) & (b).

Uniform convergence holds if we work with continuity instead of measurability. The theorem below requires that $f: X \to Y$ is continuous and that X is compact (closed and bounded) in R^n. The theorem shows that in principle a sigmoidal fuzzy system with finite fuzzy rules can approximate any continuous function to any degree of accuracy.

Theorem: A sigmoidal fuzzy system **F** uniformly approximates $f: X \to Y$ if X is compact and f is continuous.

Proof:

We pick any small constant $\varepsilon > 0$. Our aim is to show that $|F(x)-f(x)| < \varepsilon$ for all $x \in X$. given that X is a compact subset of R^n. F(x) is the centroidal output of the sigmoidal fuzzy system F in (4). Since f is continuous on the compact space X, f is uniformly continuous. So there is a fixed $\delta>0$ such that, for all x and z in X, $|f(x) - f(z)| < \frac{\varepsilon}{4}$ if $|x - z| < \delta$. We construct a set of open cubes $M_1, \ldots, M_m$, that cover X and that have ordered overlap in their n coordinates so that each cube corner lies at the midpoint c_j of its neighbors M_j.

We pick the symmetric output fuzzy sets B_j centered on $f(c_j)$. So the centroid of B_j is $f(c_j)$. Choose $u \in X$ such that u lies in at most 2^n overlapping open cubes M_j. Pick any w in the same set of cubes.

If $u \in M_j$ and $w \in M_k$, then for all $u \in M_j \cap M_k$ we have

$|u - v| < \delta$ and $|u - w| < \delta$.

Then by the uniform continuity of f, it follows that

$$|f(u)-f(w)| \le |f(u)-f(v)| + |f(v)-f(w)| < \frac{\varepsilon}{2}.$$

So for cube centers c_j and c_k, $|f(c_j)-f(c_k)| < \frac{\varepsilon}{2}$.

Let $x \in X$. Then x too lies in at most 2^n open cubes with centers c_j and $|f(c_j) - f(x)| < \frac{\varepsilon}{2}$.

Along the kth coordinate of the range space R^p the kth component of the additive system centroid F(x) lies between the kth components of the centroids of the B_j sets. So, since $|f(c_j) - f(c_k)| < \frac{\varepsilon}{2}$ for all $f(c_j)$,

$$|F(x)-f(c_j)| < \frac{\varepsilon}{2}.$$

It follows that

$$|F(x)-f(x)| \le |F(x)-f(c_j)| + |f(c_j)-f(x)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

This proves the theorem.

V.CONCLUSION

A fuzzy system or approximator reduces to a graph cover with local averaging which need be unique. An additive fuzzy system with Gaussian sets reduces to a radial basis network [3] and that too is but one of many graph coverings. The "fuzziness" or multivalence of sets comes into play when patches or output sets overlap. Nonfuzzy sets can also weigh or average the overlap. A fuzzy system is unique in the sense that it ties vague words like "small" and "medium" to the math of curves and fit vectors (points in unit cubes). So it ties natural language and commonsense rules to state-space geometry. But the "fuzzy" sets are equivalent to random sets or loci of two-point conditional probabilities. The approximation of sigmoidal fuzzy systems lies more in their model freedom than in their fuzzy interpretation. Nonfuzzy sets and rule patches also lead to model-free universal approximators.

REFERENCES

[1] J. A. Dickerson and B. Kosko, "Fuzzy Function approximation with supervised ellipsoidal learning," in *Proc. World Congress on Neural Netw. (INNS WCNN-93)*. vol. 2, July 1993, pp. 9-17.

[2] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. Orlando, FL: Academic Press, 1980.

[3] E. Hartman, J. D. Keeler, and J. Kowalski, "Layered neural networks with gaussian hidden units as universal approximators," *Neural Compu-tat.*, vol. 2, pp. 210-215, 1990.

[4] K. Hornik, M. Stinchcombe,and H. White, "Multilayerfeedforward networks are universal approximators,"*Neural Netw.*, vol. 2, pp. 359-366, 1989.

[5] B. Kosko, "Fuzzy knowledge combination," *Int. J. Intell. Syst.*, vol. 1, pp. 293-320, 1986.

[6] B. Kosko, 'Stochastic competitive learning,'' IEEE Trans. *Neural Netw.*, vol. *2, no.* 5, pp. 522-529, Sept. 1991.

[7] B. Kosko, Neural *Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Englewood Cliffs, NJ: prentice Hall, 1991.

[8] C. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison- Wesley, 1989.

[9] W. Rudin, *Functional Analysis*, New York: McGraw- Hill, 1973.

[10] W. Rudin, *Real and Complex Analysis*, second ed. New York: ZcGraw- Hill, 1974.

[11] L. Wang and J. M. Mendel, "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning," IEEE *Trans. Neural Netw.*, vol. 3, no. 5, pp. 807-814, Sept. 1992.

[12] L. A. Zadeh, "Fuzzy sets," *Inform. Contr.*, vol. 8, pp. 338-353, 1965.