# A Hierarchical Fuzzy Rule-based Learning System based on an Information Theoretic Approach

**Antony Waldock**
Advanced Technology Centre
BAE SYSTEMS
Bristol, BS34 7QW, UK
Antony.Waldock@baesystems.com

**Brian Carse**
Intelligent Autonomous Systems Laboratory
University of the West of England
Bristol, BS16 1QY, UK
Brian.Carse@uwe.ac.uk

**Chris Melhuish**
Intelligent Autonomous Systems Laboratory
University of the West of England
Bristol, BS16 1QY, UK
Chris.Melhuish@uwe.ac.uk

## Abstract

This paper proposes a new novel method for the online construction of a Hierarchical Fuzzy Rule Based System (FRBS) to accurately model a function while retaining a level of human interpretability. The algorithm uses an information theoretic approach to limit the amount of uncertainty within each decision and to determine when a rule does not effectively model the underlying decision space. Experimental results are provided which compare the performance of the proposed system with existing approaches.

**Keywords:** Fuzzy systems, machine learning, information theory.

## 1 Introduction and Related Work

The linguistic symbols in a FRBS facilitate the modelling of complex functions while retaining a degree of human interpretability. However, a model clearly interpretable by humans often leads to a reduction in accuracy. This reduction stems from the constrained structure of the linguistic symbols considered [4]. A new variant of a FRBS has been developed in an attempt to improve the accuracy of the approximation and is referred to as an Approximate FRBS [8] (also known as “non-grid-oriented” or “free semantics” FRBS [1]). An approximate FRBS increases the accuracy of the model by manipulating fuzzy sets directly instead of the linguistic symbols. By allowing the fuzzy sets to vary in number, size and position, they can be mapped directly to the data giving significant improvements [1, 6]. At the end of rule generation, the fuzzy variables are semantically free and tend to be excessively specialised. This limits the human interpretability of an approximate FRBS. However, it must be noted they often remain more interpretable than other models like neural networks.

The Hierarchical Fuzzy Rule Based System (HFRBS) has been developed, in recent years, as an attempt to improve accuracy while maintaining interpretability [7]. A HFRBS divides the input space into a fixed number of linguistic symbols each corresponding to a natural language meaning e.g. very small, small etc. Training data is used to automatically generate rules as in a standard FRBS. The key extension of a HFRBS is the use of an *expansion policy* to determine inaccurate areas of the decision space and the corresponding rules. When an inaccurate rule is identified, it is specialised into a set of smaller rules. This involves partitioning the rule’s decision space into smaller areas each represented by a separate rule. This process of specialisation continues until a desired level of accuracy is satisfied. This concept of increasing the granularity to fit the underlying decision space has been used in classifier systems for a number of years [10].

This type of linguistic symbol expansion increases the accuracy depending on the complexity of the data modelled and is commonly referred to as "ad-hoc data driven learning" [1]. The expansion can be controlled depending on the accuracy desired, human interpretability required and the complexity of the function to be approximated. Obviously some of these are mutually exclusive, for example, high human interpretability on a complex, accurate model may be impossible.

Holve outlines a method for specialisation by carefully pre-processing training data such that, when a conflict is encountered, the linguistic symbol is expanded [9]. Although demonstrated to approximate complex functions, pre-processing limits its use to applications where all the training data is available and certain. Cordón, Herrera and Zwir [7] outline a hierarchical FRBS that uses expansion techniques to specialise linguistic symbols with a large degree of error. The error of each rule is calculated by the percentage of the Mean Square Error (MSE) associated with the rule over the entire MSE from the entire training set. A rule with bad performance is determined by comparing this error with a tuneable parameter α, which dictates the rate of expansion and hence the accuracy of the function approximated. This method relies on the complete training set and test set being available during the expansion phase. For an online application, like learning behaviour in a mobile robot, this is not possible. The next section outlines the proposed algorithm with a new method of expansion based on Information Theory.

## 2 The Information Theoretic Hierarchical Fuzzy Associative Memory (IT-HFAM)

The rule generation algorithm proposed in the current contribution uses a similar method of hierarchical specialisation, but the expansion policy is determined by the amount of uncertainty of the decision within the rule. Information Theory, developed by Shannon [11], was initially concerned with modelling the efficiency of communication systems but has been applied to a multitude of other research areas including decision making and fuzzy logic systems [2]. Suppose we have a set of possible events whose probabilities of occurrence are $p_1, p_2 ... p_n$. These probabilities are known but that is all we know concerning the event. Shannon defines the amount of 'choice' involved or the uncertainty of the outcome as Entropy [11]. The entropy, $H(p_1, p_2 ... p_n)$ of an event with $n$ possible outcomes is defined by Shannon as

$$H(p_1, p_2 .. p_n) = \sum_{i=1}^{n} (p_i \times \log(p_i))$$

This measure of information can be used to calculate the uncertainty within each rule [12] and to determine the expansion of a hierarchical FRBS.

A FRBS can be viewed as a Fuzzy Associative Memory (FAM) where the linguistic symbols $M_i$, for each input $n$, produce an $n$-dimensional decision space partitioned by an $i$-dimensional grid. Each cell represents an IF-THEN rule with $n$ linguistic inputs corresponding to a single output linguistic symbol [9]. The proposed Information Theoretic Hierarchical Fuzzy Associative Memory (IT-HFAM) extends the traditional FAM approach by including an applicability distribution over all the output linguistic symbols. Figure 1 displays a FAM partitioned into four cells (rules) each with an applicability distribution over all possible output symbols.

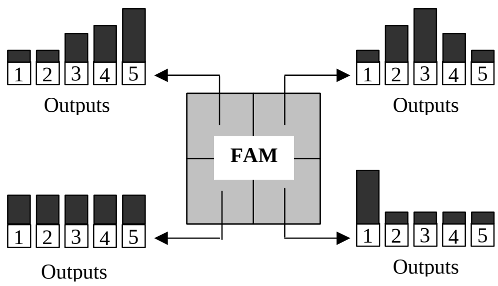


Figure 1: FAM with Output Applicability Distributions

Using Shannon's measure of uncertainty on the output applicability distribution gives the amount of choice experienced by this cell. If the applicability distribution is flat i.e. $A_i$ are equal, (Figure 1: Bottom Left Rule) then the uncertainty of the cell is at a maximum. This indicates that the cell at this granularity cannot effectively model the decision space it represents. In order to model this area of the decision space more effectively it is necessary to divide the cell into a number of smaller cells.

Training of an IT-HFAM consists of updating the applicability distribution when exposed to a new training pattern. $T_p$. Each pattern is a $(n+1)$ dimensional vector, $T_p = <x_1, x_2, x_3 \ldots x_n, y>$ where $n$ is the number of inputs; $x$ is the current input values and $y$ the target output value. The algorithm is trained on all patterns until the training set has been exhausted. The algorithm is trained by first identifying all the cells with an activation greater than zero when presented with the inputs $(x_1 .. x_n)$ from the training pattern. The applicability distribution for each active cell is updated depending on the target value $y$. For each output membership function, $M_j^y$, the applicability of the output symbol $a(M_j^y)$ is increased by the level of activation for the target value *y* (see Figure 2).

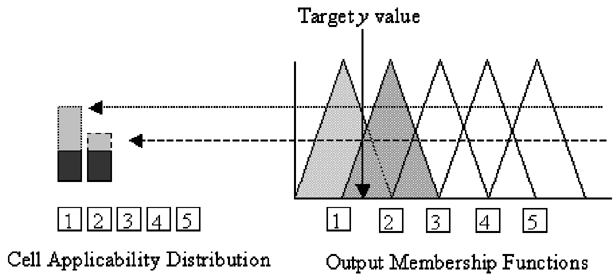


Figure 2: Cell applicability update procedure

Centre of Sums is used on the applicability distribution to calculate the current output symbol for this cell in the FAM. After each training pattern, the entropy (uncertainty) of each cell is compared with a tuneable parameter $E_{\max}$, which represents the maximum amount of uncertainty tolerated within each cell. If the uncertainty of the cell exceeds $E_{\max}$ and the applicability distribution covers more than two consecutive output symbols then the cell is divided into four smaller specialised cells (assuming the FAM consists of only two dimensions – see Figure 3).

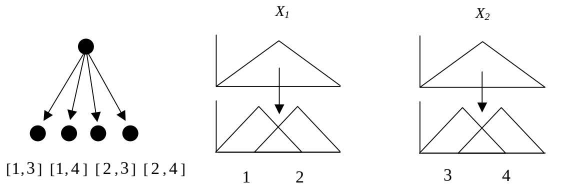


Figure 3: Membership Function Specialisation

The cell expansion is halted when only two consecutive output symbols are active in the applicability distribution. A final stage of rule reduction is applied which merges neighbouring cells with the same output fuzzy set. Rule reduction can be performed during the training phase or once after training has been complete as in the experiment in Section 3.

## 3 Experimental Results - Modelling of an Intermediate Complexity Function

This section describes the performance of the IT-HFAM algorithm, on an intermediate complexity function, compared with existing function approximators. Within the experiment below, the accuracy is determined using the Root Mean Squared Error (RMSE) equation using a test set of 100 evenly distributed points over the input space.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}((T_i - R_i)^2)}{n}}$$

Cordón, Herrera and Zwir [7,8] demonstrate the ability to train a hierarchical FRBS using a Mean Squared Error (MSE) expansion policy on an intermediate complexity function defined as:

$$f(x_1, x_2) = e^{x_1} \times \sin^2(x_2) + e^{x_2} \times \sin^2(x_1)$$

$$x_1, x_2 \in [-8,8],\ f(x_1, x_2) \in [0,5836]$$

Although Cordón et al. produce accurate approximations; the training method dictates that the entire training set is available during rule generation, which is undesirable for the type of online applications in which we are particularly interested. The IT-HFAM algorithm proposed here does not aim to improve on the accuracy of their results but demonstrate that an online solution based on information theory can achieve comparative results.

The algorithm was initialised in two ways: one with seven evenly distributed membership functions (IT-HFAM7) and the other with two user defined membership functions (IT-HFAM2) over the input space. IT-HFAM2 was initialised with 4 rules arranged such that a single rule covered the flat area of the decision space and 3 covering the raised areas on the edges. User defined membership functions allow prior knowledge about the function to direct the learning process. The system was trained using a training set of 1156 training patterns, which were evenly distributed over the function, compared to Cordón's 1089 training patterns. After exposure to a training pattern, each rule updates its applicability distribution and determines whether specialisation is required. The RMSE was calculated after exposure to the complete training set and then after a second iteration. Both had seven evenly distributed membership functions over the output. The results of the IT-HFAM algorithm using three different entropy limits are compared in Tables 1a and 1b with a comparison to existing function approximators in Table 2. For each experiment, the number of cells (rules) after training ($C_t$), the number of cells after reduction ($C_r$) and the final accuracy before ($RMSE_t$) and after reduction ($RMSE_r$) are displayed.

Table 1a – IT-HFAM Results

| Experiment | 1 Training Set Iteration | | | |
|---|---|---|---|---|
| | **$C_t$** | **$C_r$** | **$RMSE_t$** | **$RMSE_r$** |
| IT-HFAM2 – 90% | 7 | 4 | 0.0951 | 0.0941 |
| IT-HFAM2– 70% | 13 | 7 | 0.0981 | 0.0913 |
| IT-HFAM2– 30 % | 3181 | 233 | 0.0596 | 0.0768 |
| IT-HFAM7 – 90% | 49 | 3 | 0.1014 | 0.1036 |
| IT-HFAM7– 55% | 247 | 55 | 0.0951 | 0.0857 |
| IT-HFAM7– 45 % | 925 | 153 | 0.0717 | 0.0863 |

Table 1b – IT-HFAM Results

| Experiment | 2 Training Set Iterations | | | |
|---|---|---|---|---|
| | **$C_t$** | **$C_r$** | **$RMSE_t$** | **$RMSE_r$** |
| IT-HFAM2-90% | 16 | **7** | 0.0898 | **0.0880** |
| IT-HFAM2-70% | 43 | 14 | 0.0778 | 0.0771 |
| IT-HFAM2-30% | 11014 | 274 | 0.0375 | 0.0441 |
| IT-HFAM7-90% | 49 | 3 | 0.1014 | 0.1036 |
| IT-HFAM7-55% | 1003 | 156 | 0.0545 | 0.0578 |
| IT-HFAM7-45% | 3952 | **336** | 0.0242 | **0.0291** |

Table 2 - Comparison with existing function approximators

| Experiment | No Rules | RMSE |
|---|---|---|
| S-WCA fixed [3] | 9 | 0.0868 |
| S-WCA hierarchical [7,8] | 316 | 0.0134 |
| FCM [5] | 6 | 0.1124 |
| FCM [7,8] | 9 | 0.0403 |
| IT-HFAM | 7 | 0.0880 |
| IT-HFAM | 336 | 0.0291 |

The results shown in Table 2 demonstrate that the on-line IT-HFAM performance is comparable to the other off-line methods. Tables 1a and 1b demonstrate how adjusting the amount of uncertainty tolerated within the decision acts as a trade off between the number of rules generated and the accuracy of the model. Figure 4 demonstrates this further by showing how the accuracy of the model increases when the amount of uncertainty tolerated is decreased.

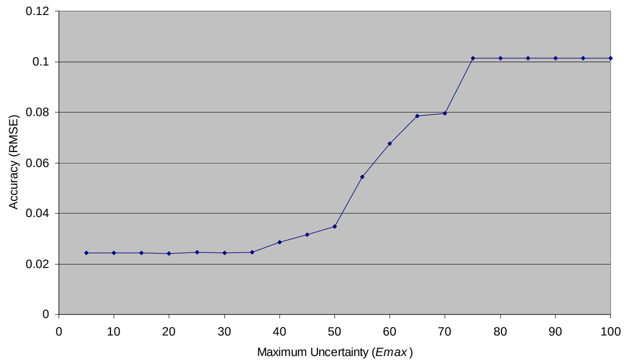


Figure 4: Accuracy vs Uncertainty tolerated

## 4 Conclusions and Further Work

A new novel method for learning hierarchical fuzzy rule-bases based on an information theoretic approach (IT-HFAM) has been proposed. The method has been evaluated and compared with existing approaches published elsewhere. IT-HFAM, devised for on-line learning, compares reasonably well in terms of accuracy and number of learned rules with several off-line approaches, which require the complete training set at the outset. The next stage in this work is to evaluate the performance of IT-HFAM for real-world, on-line learning applied to mobile robotics.


## Acknowledgements

This work was funded by the Advanced Technology Centre Filton, BAE SYSTEMS Ltd.


## References


[1] R. Alcala, J. Casillas, O. Cordón, F Herrera. Building Fuzzy Graphs: Features and Taxonomy of Learning for Non-grid-oriented Fuzzy Rule-based Systems, Journal of Intelligent and Fuzzy Systems 11, 2001, pp. 99-119.

[2] S. Al-sharhan, F. Karray. W. Gueaieb and O. Basir (2001). Fuzzy Entropy: a Brief Survey. The 10th IEEE International Fuzzy Systems Conference. Volume 3. pp 1135-1139.

[3] A. Bardossy, L. Duckstein (1995) Fuzzy Rule-Based Modelling with Application to Geophysical, Biological and Engineering Systems, CRC Press.

[4] A. Bastian (1994), How to handle the flexibility of linguistic variables with applications, Intl. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 2:4 , pp463-484

[5] J. C. Bezdek, S. Pal (Eds) (1992). Fuzzy Models for Pattern Recognition: Methods that search for structures in Data. IEEE Press.

[6] B. Carse, T.C. Fogarty and A. Munro(1996). Evolutionary Learning of Fuzzy Rule Based Controllers using Genetic Algorithms. Fuzzy Sets and Systems 80 pp. 273-293.

[7] O. Cordón, F. Herrera, I. Zwir (2001). Fuzzy Modelling by Hierarchically built Fuzzy Rule Bases. International Journal of Approximate Reasoning Vol. 27 pp. 61-93.

[8] O. Cordón, F. Herrera, F. Hoffmann & L. Magdalena(2001). Genetic Fuzzy Systems, Evolutionary Tuning and Learning of Fuzzy Knowledge Bases. World Scientific, ISBN 981-02-4016-3.

[9] R. Holve(1998). Investigation of Automatic Rule Generation for Hierarchical Fuzzy Systems. Proc of FUZZ IEEE at 1998 IEEE World Congress On Computational Intelligence, May 4-9, 1998, Anchorage, Alaska pp973-978.

[10] C. Melhuish and T. Fogarty. Applying A Restricted Mating Policy To Determine State Space Niches Using Immediate and Delayed Reinforcement. Evolutionary Computing. AISB Workshop. Leeds UK 94 p224-237

[11] C. E. Shannon(1948), A mathematical theory of communication, Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October.

[12] L. Zadeh(1965). Probability measures of fuzzy events. J. Math. Anal. Appl., 23:421-427.