Gradient-based Parameter Optimisation of FRI "FIVE"

Zoltán Krizsán, Szilveszter Kovács

Department of Information Technology, University of Miskolc
Miskolc-Egyetemváros, Miskolc, H-3515, Hungary
krizsan@iit.uni-miskolc.hu
http://www.iit.uni-miskolc.hu/krizsan
szkovacs@iit.uni-miskolc.hu
http://www.iit.uni-miskolc.hu/~szkovacs

Abstract: The main contribution of this paper is the extension of an existing Fuzzy Rule Interpolation (FRI) method by a gradient-based consequent optimisation. The targeted FRI method is an application oriented approach called "FIVE" (Fuzzy Rule Interpolation based on the Vague Environment of the Fuzzy Rule Base). The goal of the consequent optimisation is the rule base fine tuning in the case if there are input-output sample data of the modelled system exists.

Keywords: Fuzzy Rule Interpolation, FRI, FIVE, gradient based rule optimisation

1 Inroduction

Fuzzy Rule Interpolation (FRI) based fuzzy systems – whose rule base is spare – do not have rules for all the possible observations, in other words it could exists one (or more) observation(s) that does not lead to an interpretable conclusion applying the classical fuzzy reasoning methods. In this case the conventional fuzzy inference (Zadeh, Mamdani, Larsen, or Takagi-Sugeno) cannot be used. There are lots of FRI methods can be found in the literature, and every method has its advantage. Some of them are very precise, some of them are less precise but its computation time is better.

An application oriented aspect of the FRI emerges in the concept of "FIVE". The fuzzy reasoning method "FIVE" (Fuzzy Interpolation based on Vague Environment, originally introduced in [2], [3] and [4]) was developed to fit the speed requirements of direct fuzzy control, where the conclusions of the fuzzy controller are applied directly as control actions in a real-time system (see e.g. a downloadable and runable code of a real-time vehicle path tracking and collision avoidance control at [7]).

In the followings first the FRI method FIVE will be introduced in more details, then the paper will suggest a gradient-based optimisation method (steepest descent) for the automatic consequent optimisation of the FIVE rule base.

2 The Concept of "FIVE"

The FIVE FRI method is based on the concept of the vague environment [1]. Applying the idea of the vague environment the linguistic terms of the fuzzy partitions can be described by scaling functions [1] and the fuzzy reasoning itself can be replaced by classical interpolation. The concept of a vague environment is based on the similarity or indistinguishability of the considered elements. Two values in a vague environment are ε-distinguishable if their distance is greater than ε. The distances in a vague environment are weighted distances. The weighting factor or function is called *scaling function (factor)* [1].

Two values in the vague environment *X* are *ε-indistinguishable* if

$$\varepsilon \geq \delta_s(x_1, x_2) = \left| \int_{x_2}^{x_1} s(x)dx \right| , \tag{1}$$

where $\delta_s(x_1, x_2)$ is the scaled distance of the values x_1, x_2 and $s(x)$ is the scaling function on *X*.

For finding connections between fuzzy sets and a vague environment the membership function $\mu_A(x)$ can be introduced as indicating level of similarity of *x* to a specific element *a* that is a representative or prototypical element of the fuzzy set $\mu_A(x)$, or, equivalently, as the degree to which *x* is indistinguishable from *a* (2) [1]. The α-cuts of the fuzzy set $\mu_A(x)$ are the sets which contain the elements that are (1−α)-indistinguishable from *a* (see Fig. 1):

$$1-\alpha \geq \delta_s(\mathrm{a,b}) \ , \quad \mu_A(x) = 1 - \min\{\delta_s(\mathrm{a,b}),1\} = 1 - \min\left\{ \left| \int_{\mathrm{a}}^{\mathrm{b}} s(x)dx \right|, 1 \right\} . \tag{2}$$

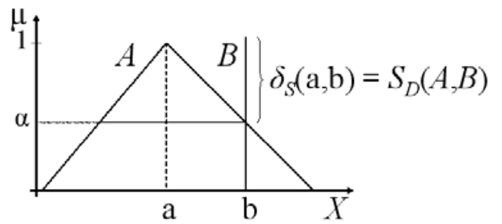

Figure 1
The α-cuts of $\mu_A(x)$ contain the elements that are (1−α)-indistinguishable from *a*

In this case (see **Error! Reference source not found.**), the scaled distance of points a and b ($\delta_s(\mathrm{a,b})$) is the *Disconsistency Measure* (S_D) (mentioned and studied among other distance measures in [6] by Turksen *et al.*) of the fuzzy sets *A* and *B* (where *B* is a singleton):

$$S_D(A,B) = 1 - \sup_{x\in X} \mu_{A\cap B}(x) = \delta_s(\mathrm{a,b}) \text{ if } \delta_s(\mathrm{a,b}) \in [0,1] \;, \tag{3}$$

where $A \cap B$ notes the min t-norm, $\mu_{A\cap B}(x) = \min[\mu_A(x), \mu_B(x)]$ $\forall x \in X$.

Taking into account the most common way of building a traditional fuzzy logic controller, where the first step is defining the fuzzy partitions on the antecedent and consequent universes by setting up the linguistic terms and then based on these terms building up the fuzzy rule base, the concept of vague environment [1] is straightforward. The goal of the fuzzy partitions is to define indistinguishability, or vagueness on the different regions of the input, output universes.

The vague environment is characterised by its scaling function. For generating a vague environment of a fuzzy partition an appropriate scaling function is needed, which describes the shapes of all the terms in the fuzzy partition. Generally a fuzzy partition can not be characterised by a single vague environment, so the question is how to describe all fuzzy sets of the fuzzy partition with one "universal" scaling function. For this task the concept of an *approximate scaling function*, as an approximation of the scaling functions describing the terms of the fuzzy partition separately is proposed in [2], [3], [4].

3 Shepard Interpolation for "FIVE"

The main idea of the FRI method "FIVE" (Fuzzy Interpolation based on Vague Environment) can be summarised in the followings:

a) If the vague environment of a fuzzy partition (the scaling function or at least the approximate scaling function) exists, the member sets of the fuzzy partition can be characterised by points in that vague environment. (These points are indicating the cores of the fuzzy terms, while the membership functions are described by the scaling function itself.)

b) If all the vague environments of the antecedent and consequent universes of the fuzzy rule base exist, all the primary fuzzy sets (linguistic terms) compounding the fuzzy rule base can be characterised by points in their vague environment. Therefore the fuzzy rules (built-up from the primary fuzzy sets) can be characterised by points in the vague environment of the fuzzy rule base too. In this case the approximate fuzzy reasoning can be handled as a classical interpolation task.

c) Applying the concept of vague environments (the distances of points are weighted distances), any crisp interpolation, extrapolation, or regression method can be adapted very simply for approximate fuzzy reasoning [2], [3] and [4].

Because of its simple multidimensional applicability, for interpolation-based fuzzy reasoning in this paper the adaptation of the *Shepard operator* based interpolation (first introduced in [5]) is suggested. The Shepard interpolation method for arbitrarily placed bivariate data was introduced as follows [5]:

$$S_0(f,x,y)=\begin{cases} f_k & \text{if } (x,y)=(x_k,y_k) \text{ for some } k, \\ \left(\sum_{k=0}^{n} f(x_k,y_k)/d_k^{\lambda}\right) \Big/ \left(\sum_{k=0}^{n} 1/d_k^{\lambda}\right) & \text{otherwise,} \end{cases} \quad (4)$$

where measurement points x_k, y_k ($k \in [0,n]$) are irregularly spaced on the domain of $f \in \Re^2 \to \Re$, $\lambda > 0$, and $d_k = \left[(x-x_k)^2 + (y-y_k)^2\right]^{\frac{1}{2}}$. This function can be typically used when a surface model is required to interpolate scattered spatial measurements.

The adaptation of the Shepard interpolation method for interpolation-based fuzzy reasoning in the vague environment of the fuzzy rule base is straightforward by substituting the Euclidian distances d_k by the scaled distances $\delta_{s,k}$:

$$\delta_{s,k} = \delta_s(\mathbf{a}_k,\mathbf{x}) = \left[\sum_{i=1}^{m}\left(\int_{\mathrm{a}_{k,i}}^{\mathrm{x}_i} s_{\mathrm{X}_i}(\mathrm{x}_i)\,d\mathrm{x}_i\right)^2\right]^{1/2}, \quad (5)$$

where s_{X_i} is the i^{th} scaling function of the m dimensional antecedent universe, $\mathbf{x}$ is the m dimensional crisp observation and $\mathbf{a}_k$ are the cores of the m dimensional fuzzy rule antecedents $\boldsymbol{A}_k$.

Thus in case of singleton rule consequents the fuzzy rules R_k has the following form:

$$\textbf{If } \mathrm{x}_1 = A_{\mathrm{k},1} \textbf{ And } \mathrm{x}_2 = A_{\mathrm{k},2} \textbf{ And } \ldots \textbf{ And } \mathrm{x}_m = A_{\mathrm{k},m} \textbf{ Then } \mathrm{y} = \mathrm{c_k} \quad (6)$$

by substituting (5) to (4) the conclusion of the interpolative fuzzy reasoning can be obtained as:

$$\mathrm{y}(\mathbf{x})=\begin{cases} \mathrm{c}_k & \text{if } \mathbf{x}=\mathbf{a}_k \text{ for some } k, \\ \left(\sum_{k=1}^{r} \mathrm{c}_k/\delta_{s,k}^{\lambda}\right) \Big/ \left(\sum_{k=1}^{r} 1/\delta_{s,k}^{\lambda}\right) & \text{otherwise.} \end{cases} \quad (7)$$

4 Gradient-based Consequent Optimisation

The main contribution of this paper is the suggestion of a gradient-based optimisation method (steepest descent) for the consequent optimisation of the FIVE rule base. The consequent optimisation is based on a set of sample (training) data. The goal of the optimisation method is minimising the squared error E of the training data and the fuzzy model:

$$E = \sum_{k=1}^{N} \left(y_d(x_k) - y(x_k) \right)^2 , \tag{8}$$

where $y_d(x_k)$ is the desired output of the kth training data and $y(x_k)$ is the output of the FIVE fuzzy model, N is the number of the training data.

The applied steepest descent parameter optimisation method modifies the rule consequents based on their partial derivatives to the squared error function E (8) in the following manner:

$$g(c_k) = \frac{\partial E(c_k)}{\partial c_k} = \frac{\partial E(c_k)}{\partial y(x)} \cdot \frac{\partial y(x)}{\partial c_k} \tag{9}$$

$$c_{k\,next} = c_k - \eta \cdot g(c_k), \tag{10}$$

where η is the step size of the iteration and c_{knext} is the next iteration of the kth conclusion c_k.

According to (8), (9) can be rewritten in the following form:

$$g(c_k) = -2 \cdot \left(y_d(x_k) - y(x_k) \right) \cdot \frac{\partial y(x)}{\partial c_k} \tag{11}$$

Applying the Shepard interpolation formula of FIVE (7), for the partial derivatives we get the following formulas:

$$\frac{\partial \mathrm{y}(\mathbf{x})}{\partial c_k} = \begin{cases} 1 & \text{if } \mathbf{x} = \mathbf{a}_k \text{ for some } k, \\ \left(1/\delta_{s,k}^{\lambda}\right) \Big/ \left(\sum_{k=1}^{r} 1/\delta_{s,k}^{\lambda} \right) & \text{otherwise.} \end{cases} \tag{12}$$

According (10), (11) and (12) the next iteration of the kth conclusion c_k can be calculated.

5 Application Example

The training data of the application example is a simple input-output set of randomly selected data from the y=sin(x)/x function in the domain of [-20, 20]. For demonstration purposes this domain is evenly covered by 13 single input single output fuzzy rules in the following form (for the k^{th} rule of the rule base):

$$\textbf{If } x = A_k \quad \textbf{Then } y = c_k \tag{13}$$

For the initial state of the experiment all the consequents of the fuzzy rules are set to 1 (c_k=1, $k \in [1,13]$).

The antecedents (A_k) of the fuzzy rules are fixed, and more or less evenly distributed in the domain according to the Fig. 2.

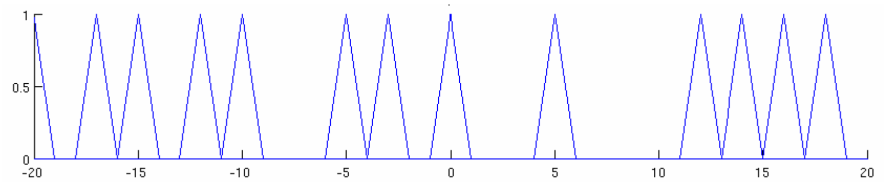

Figure 2
The fixed antecedents (A_k) of the 13 fuzzy rules

Fig. 3 introduces the values of the training data, the conclusions of the initial, and the parameter optimised fuzzy model. The change of the squared error of the training data and the fuzzy model (8) in the function of the iteration steps is demonstrated on Fig. 4.

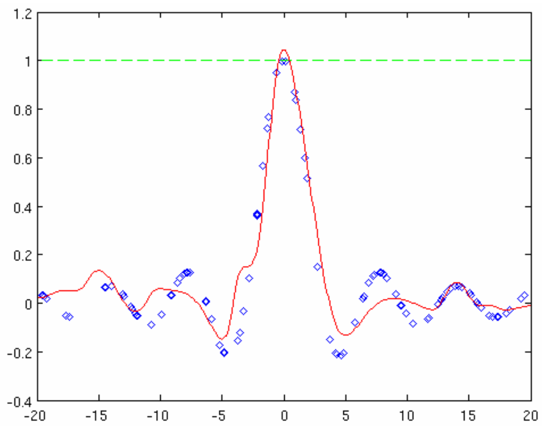

Figure 3
The training data (circles), the conclusions of the initial (dashed line), and the parameter optimised fuzzy model (continuous line)

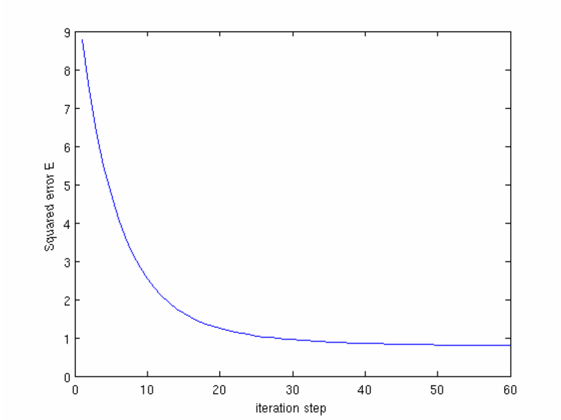

Figure 4

The change of the squared error (8) in the function of the iteration steps

Conclusions

As a first step of automatic rule base generation for FRI methods, this paper suggests a gradient-based optimisation method (steepest descent) for the consequent optimisation of the FIVE rule base. The consequent optimisation is based on a set of sample (training) data. The goal of the optimisation method is minimising the squared error of the training data and the FRI fuzzy model. In case of the method FIVE the suggested solution is rather simple and efficient, serving a good base for the further improvement of fully automatic FRI FIVE rule base generation.

Acknowledgement

This research was partly supported by the Intelligent Integrated Systems Japanese Hungarian Laboratory and the National Scientific Research Fund Grant OTKA T049499.

References

[1] Klawonn, F.: Fuzzy Sets and Vague Environments, Fuzzy Sets and Systems, 66, pp. 207-221 (1994)

[2] Kovács, Sz.: New Aspects of Interpolative Reasoning, Proceedings of the 6th. International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Granada, Spain, pp. 477-482 (1996)

[3] Kovács, Sz., Kóczy, L. T.: Approximate Fuzzy Reasoning Based on Interpolation in the Vague Environment of the Fuzzy Rule base as a Practical Alternative of the Classical CRI, Proceedings of the 7th International Fuzzy Systems Association World Congress, Prague, Czech Republic, pp. 144-149 (1997)

[4] Kovács, Sz., Kóczy, L. T.: The Use of the Concept of Vague Environment in Approximate Fuzzy Reasoning, Fuzzy Set Theory and Applications, Tatra Mountains Mathematical Publications, Mathematical Institute Slovak Academy of Sciences, Bratislava, Slovak Republic, Vol. 12. pp. 169-181 (1997)

[5] Shepard, D.: A Two Dimensional Interpolation Function for Irregularly Spaced Data, Proc. 23rd ACM Internat. Conf., pp. 517-524 (1968)

[6] Turksen, I. B., Zhong, Z.: An Approximate Analogical Reasoning Scheme based on Similarity Measures and Interval Valued Fuzzy Sets, Fuzzy Sets and Systems, 34, pp. 323-346 (1990)

[7] Some FRI applications are available at: http://www.iit.uni-miskolc.hu/~szkovacs/