# Qualitative Optimization of Fuzzy Causal Rule Bases using Fuzzy Boolean Nets♣

João Paulo Carvalho[1, 2] José Tomé[1, 2]

*1.INESC-ID, R. Alves Redol 29, 1000-029 Lisboa, Portugal*

*2.Instituto Superior Técnico, Universidade Técnica de Lisboa, 1000 Lisboa, Portugal*

*Email: joao.carvalho@inesc-id.pt, jose.tome@inesc-id.pt*

**ABSTRACT**: Fuzzy Causal Rule Bases (FCRb) are widely used and are the most important rule bases in Rule Based Fuzzy Cognitive Maps (RB-FCM). However, FCRb are subject to several restrictions that create difficulties in their creation and completion. This paper proposes a method to optimally complete Fuzzy Causal Rule bases using Fuzzy Boolean Net properties as qualitative universal approximators. Although the proposed approach focuses on FCRb, it can be generalized to any fuzzy rule base.

**Keywords:** Fuzzy Boolean Nets, Fuzzy Causal Relations, Rule base Optimization

## 1. Introduction

Rule Based Fuzzy Cognitive Maps (RB-FCM), are a qualitative approach to modeling and simulating the Dynamics of Qualitative Systems (like, for instance, Social, Economical or Political Systems) [5][6][8]. RB-FCM were developed as a tool that can be used by non-engineers and/or non-mathematicians and eliminates the need for complex mathematical knowledge when modeling qualitative dynamic systems. Fuzzy Causal Relations (FCR) were previously introduced in [1][2][7], and are the most common method to describe the relations between the entities (known as concepts) of RB-FCM [1][2][7]. FCR are represented and defined through linguistic Fuzzy Causal Rule Bases (FCRb). RB-FCM inference imposes that Fuzzy Causal Rule Bases must be complete and involve only one antecedent (multiple antecedent inference is dealt with internal RB-FCM mechanisms, like the Fuzzy Causal Accumulation operation [1][5][7]). It also imposes certain strict restrictions to the linguistic terms involved in the inference [1][2][7](section 2.). Another important characteristic of FCRb is the unusually large number of linguistic terms needed to properly represent the involved relations in typical applications (variables with 11 or 13 linguistic terms are common in FCRb.) [5][6].

FCR data is usually obtained through "far from ideal" methods, and all of the above characteristics and restrictions mean that extra-special care must be taken with FCRb construction when modeling a RB-FCM, especially when one considers that RB-FCM usually contain a large number of feedback cycles, and as a consequence, minor differences in modeling result in drastic simulation differences and conclusions [4][6][17]. Therefore, FCR data must often be optimized before it can be used on the RB-FCM. As we will see next, this optimization process differs from classical optimization problems due to several factors.

On a RB-FCM, experts usually express knowledge using just a few key rules. These key rules should contain enough knowledge for a human expert to extrapolate all the rules necessary to complete the FCRb. An optimal FCRb consists on the set of rules extrapolated by an expert from the available key rules. In FCRb such set must be complete (see section 2.) and usually has a cardinality of 11 or 13 [5][6]. Under this definition, the optimal FCRb has several characteristics that prevent the use of a classic quantitative optimization approach. The problem of obtaining the

♣ This work is partially supported by the FCT - Portuguese Foundation for Science and Technology under project POSI/SRI/47188/2002

optimal set lies in the fact that what is relatively easy for a human is not necessarily easily automated when one deals with qualitative linguistic knowledge representation. For example, a simple linear relation between two linguistic fuzzy variables cannot necessarily be represented by a quantitative linear expression. A qualitative linear relation is usually expressed by an expert using just two fuzzy "if...then" rules (just as one would only need two points to define a straight line on a quantitative problem). However, if the membership functions used to define the linguistic terms are not identical and/or are not regularly distributed over the Universe of Discourse (UoD), we are indeed in presence of a quantitative non-linear relation that possibly cannot be optimally approached using a linear function. Note that in the simplest case, where the number of antecedent and consequent linguistic terms is identical, finding the optimal linguistic FCRb is trivial for an expert (see Section 5., example 2), while a numerical approach can be far from trivial due to the non-linear constraints imposed by membership function shape, size and centre of gravity. Therefore optimal rule base completion should probably focus on alternative qualitative methods.

An additional difficulty concerning FCRb optimal completion lies in the fact there is no way to always objectively define an optimal solution, although that is possible in some particular cases (like when the expert is defining a qualitative linear relation and the number of antecedent and consequent linguistic terms is identical). In fact, experts often provide a small number of rules even when expressing relations that are far from linear. Humans can usually still complete the FCRb using common sense (a quantitative 2-dimensional analogy would be imagining a line smoothly linking a few key points), but only the original expert would know for sure if the completion is optimal. In these cases, one simply cannot define an objective function to minimize, since only the expert that gave the key rules could properly validate a result. The best solution available is to make a "smooth" qualitative interpolation using as constraints the available key rules and the linguistic term set (the 2D analogy optimal solution would be the interpolation that gives the "smoothest line" and passes by a set of predefined points). Thus, for this study, an optimal result is obtained by finding the smoothest qualitative interpolation that respects all constraints.

The problem of optimal FCRb completion can become even more complex when one considers the cases where data comes from more than one expert or from quantitative measurements. The optimization process should therefore be able to cope with the following situations:

- *Expert knowledge*
    - *Single expert case*: the expert usually expresses knowledge using just a few key rules that must be generalized to all UoD – one must find an optimal FCRb using a completion process;
    - *Multiple expert case*: different experts might have different opinions on the same problems, and, as a result, supplied key rules can be different – one must find an optimal FCRb by combining different and possibly inconsistent rules, and by rule base completion;
- *Uncertain and sparse quantitative data*: in some cases the need arises to obtain a FCRb from observations, measurements, etc. – finding an optimal rule base involves qualitative rule extraction from data followed by rule base completion;

Several methods have been proposed to address the above (or part of the above) problems. However, although those methods are valid in most fuzzy rule base completion problems, they fall short when dealing with FCR optimization for several reasons we present in the next section.

To solve the problem of optimal FCRb completion, we propose the use of Fuzzy Boolean Nets (FBN). FBN have been previously introduced as a hybrid fuzzy-neural technique where fuzziness is an emergent property that gives FBN the capability of extracting qualitative fuzzy rules from quantitative data [18][20][22]. FBN are qualitative universal approximators [19] with an excellent generalization capability that allows them to perform as qualitative interpolators. It is our contention that FBNs produce an optimal completion because they generate the "best" and most robust smooth qualitative rule base.

## 2. Fuzzy Rule Base Interpolation and Completion Methods

Methods to deal with incomplete (or sparse) rule bases can be divided into three major categories:

1) Rule interpolation;
2) Analogical inference;
3) Rule base completion.

Valerie Cross and Thomas Sudkamp work [9] is an excellent starting point and provide lots of useful references regarding this topic. The first two categories can be considered "on line" in a sense that whenever an input occurs in a region of the UoD not covered by the existing rule base, proximity and similarity to the nearest rules are used to produce an output [9][10][12][14]. On the other hand, "rule base completion" methods are "off line" methods, since additional rules are created before any inference occurs.

"On line" methods are not a valid choice to solve the problem of FCRb completion because all linguistic terms involved in FCR inference must be Interpolated Linguistic Terms (ILT), and comply with the following strict restrictions [1][2][3][5][7]:

**i.** The membership degree of all linguistic terms must be complementary, i.e., its sum must be 1 in every point of the variable UoD (X):

$$\forall x \in \mathrm{X}, \forall (A_0, A_1, ..., A_n) \in \mathrm{F}(\mathrm{X}), \sum_{\mathrm{i}=0}^{\mathrm{n}} \mu_{A_i}(x) = 1 \quad (1)$$

**ii.** All linguistic terms must have the same basic shape (trapezoidal, S, etc.), and their membership functions must cross with their neighbours when $\mu$=0.5.

**iii.** The inference method must preserve both the shape and the centroid's x-coordinate of the consequent linguistic term; the Max-Dot method is an example of an adequate method.

**iv.** The fuzzy sets that result from the inference of the rule base must be summed. As a result one obtains a single fuzzy set, which we will call U.

An ILT, Interpolated Linguistic Term [3], is a fuzzy set that is univocally related with the active consequents of a rule based fuzzy inference. Given U, obtained respecting restrictions i. to iv., we call $ILT_U$ (the Interpolated Linguistic Term of $U$), to the fuzzy set that respects the following conditions:

**v.** $ILT_U$ and the term set of the fuzzy variable where $U$ is defined must have the same shape.

**vi.** The x-coordinate of the centroid of $ILT_U$ and the x-coordinate of $U$ must be the same:

$$\mathrm{xC}_{ILT_U} = \mathrm{xC}_U \Leftrightarrow \left( \frac{\int_x \mu_{ILT_U}(x) \cdot x \; dx}{\int_x \mu_{ILT_U}(x) \; dx} \right) = \left( \frac{\int_x \mu_U(x) \cdot x \; dx}{\int_x \mu_U(x) \; dx} \right) \quad (2)$$

**vii.** $U$ and $ILT_U$ must have the same Area:

$$\text{Area}_{ILT_U} = \text{Area}_U \Leftrightarrow \int_X \mu_{ILT_U}(x)\ dx = \int_X \mu_U(x)\ dx \tag{3}$$

**viii.** $ILT_U$ is normal, i.e.:

$$\{\exists x \in X \mid \mu_{ILT_U}(x) = 1\} \Leftrightarrow ILT_{U1} \neq \varnothing \Leftrightarrow \text{xtop}_{ILT_U} > 0, \tag{4}$$

where $ILT_{U1}$ represents the α-cut of $ILT_U$ for α =1 and xtop is the size of $ILT_{U1}$.

**ix.** If $A$ and $B$ are the terms involved in the inference of $U$, then the size of $ILT_{U1}$, $\text{xtop}_{ILTU}$, is a function of $A$ and $B$'s xtop and of $A$, B and $U$'s xC:

$$\text{xtop}_{ILT_U} = \min\{\text{xtop}_A, \text{xtop}_B\} + \left| \frac{\text{xC}_U - \text{xC}_A}{\text{xC}_B - \text{xC}_A} \times (\text{xtop}_A - \text{xtop}_B) \right| \tag{5}$$

**x.** If $A$ and $B$ are the terms involved in the inference of $U$, then the size of the inner base of $ILT_U$, $\text{bi}_{ILTU}$, is a function of $A$ and $B$'s bi and of $A$, B and $U$'s xC:

$$\text{bi}_{ILT_U} = \min\{\text{bi}_A, \text{bi}_B\} + \left| \frac{\text{xC}_U - \text{xC}_A}{\text{xC}_B - \text{xC}_A} \times (\text{bi}_A - \text{bi}_B) \right| \tag{6}$$

Since "On line" completion techniques do not comply with the above conditions, one must resort to "off line" rule base completion methods.

"Off line" completion techniques can be divided into two categories: those that do not require predetermined fuzzy partitions (linguistic terms) of the input and output domain [23], and those that require them [15][16][24]. Once again, the strict linguistic term restrictions of FCR prevent the use of the former techniques. Therefore, FCRb completion is confined to one of the variations of the latter technique.

In this technique one must add a rule

"If $X$ is $A_i$ Then $Z$ is $C_j$ "

for each antecedent linguistic term $A_i$ without reference in the rule base[1]. A scalar value $z_j$ must be generated using either training data or nearby rules. The selected consequent linguistic term $C_j$ will be the one where $z_j$ has maximal membership. The variations differ in the way how $z_j$ is generated:

- One can use available training data to learn and generate the rule [16][24];
- One can use the neighboring rules $A_{i-1}$ and $A_{i+1}$ to obtain $z_j$ through Region Growing [16];
- One can use all rules in the rule base to obtain $z_j$ through Weighted Averaging [16];
- One can obtain $z_j$ through Interpolation by Similarity of available rules [15];

Although these approaches are valid in most problems, they fall short when dealing with the FCR optimization problem for several reasons:

- Automatic rule extraction from quantitative data is usually based on TPE systems [16][24] that are incompatible with the FCR linguistic term set restrictions [1][2][7], and that need an unusually large number of training examples in order to produce a complete rule base containing fuzzy variables with 11 or 13 linguistic terms [16]. This is a serious problem when dealing with qualitative data from real-world experts, and in the end one often has to resort to the other completion techniques. Therefore this technique can be used to automatically create rule bases from quantitative data, but is not very adequate to produce

[1] Note that FCR always have a single antecedent.

complete FCRb;

- In general, completion methods do not produce useful results when the data is too sparse, even in linear problems (see section 5). Unfortunately, very sparse raw FCR are pretty common (see section 5);
- Region growing techniques simply do not produce good results when completing FCRb obtained from expert knowledge (see section 5). This is due to the fact that completion is too "local", and rules with a single neighbor maintain its neighbor consequent, which is, as we will show, an undesirable behavior in raw FCRb optimization;
- On the other hand, weighted averaging produces undesirable and uncontrollable results in several situations [16](section 5.). This due to the fact that all rules are considered in the completion process (too much global interference);
- Region growing and Weighted Averaging can be considered the extreme cases of a technique known as Interpolation by Similarity. This technique can be "tailored" to produce much better results than the previous ones according to each case. However, this technique is not automatic or "user transparent" and often needs strong parameterization before can be applied to each particular case. Therefore it does not comply with RB-FCM's philosophy of accessibility for users without advanced mathematical knowledge;
- Finally, these methods do not provide mechanisms to deal automatically with the problem of inconsistent opinions from several experts. Therefore other methods must be used to complement them.

## 3. Fuzzy Boolean Nets

Natural or Biological neural systems have a certain number of features that leads to their learning capability when exposed to sets of experiments from the real outside world. They also have the capability to use the learnt knowledge to perform reasoning in an approximate way. Fuzzy Boolean Nets (FBN) were developed with the goal of exhibiting these properties [19][20]. FBN can be considered a neural fuzzy model where the fuzziness is an inherent emerging property; in other known models, fuzziness is artificially introduced on neural nets, or neural components are inserted on the fuzzy systems.

In FBN, neurons are grouped into areas. Each area can be associated with a given variable, or concept. Meshes of weightless connections between antecedent neuron outputs and consequent neuron inputs are used to perform *If…Then* inference between areas. Neurons are binary, and the meshes are formed by individual random connections (just like in nature). Each consequent neuron contains $m$ inputs for each antecedent area, that is a total of $N.m$ inputs, and up to $(m+1)^N$ internal unitary memories (called FF [2]), where $N$ is the number of antecedents. This number corresponds to maximum granularity, and can be reduced. In this case, each one of the FF's is addressed by a unique joint count of activated inputs on the $m$ inputs and $N$ antecedents. It is considered that each neuron's internal unitary memories (FF) can also have a third state meaning "not taught". As in nature, the model is robust in the sense that it is immune to individual neuron or connection errors (which is not the case of other models, such as the classic artificial neural net) and presents good generalization capabilities.

The "value" of each concept, when stimulated, is given by the activation ratio of its associated area (which is given by the relation between active - output "1" - neurons and the total number of neurons).

[2] FF stands for Flip-Flop, which is a 1 bit memory digital circuit.

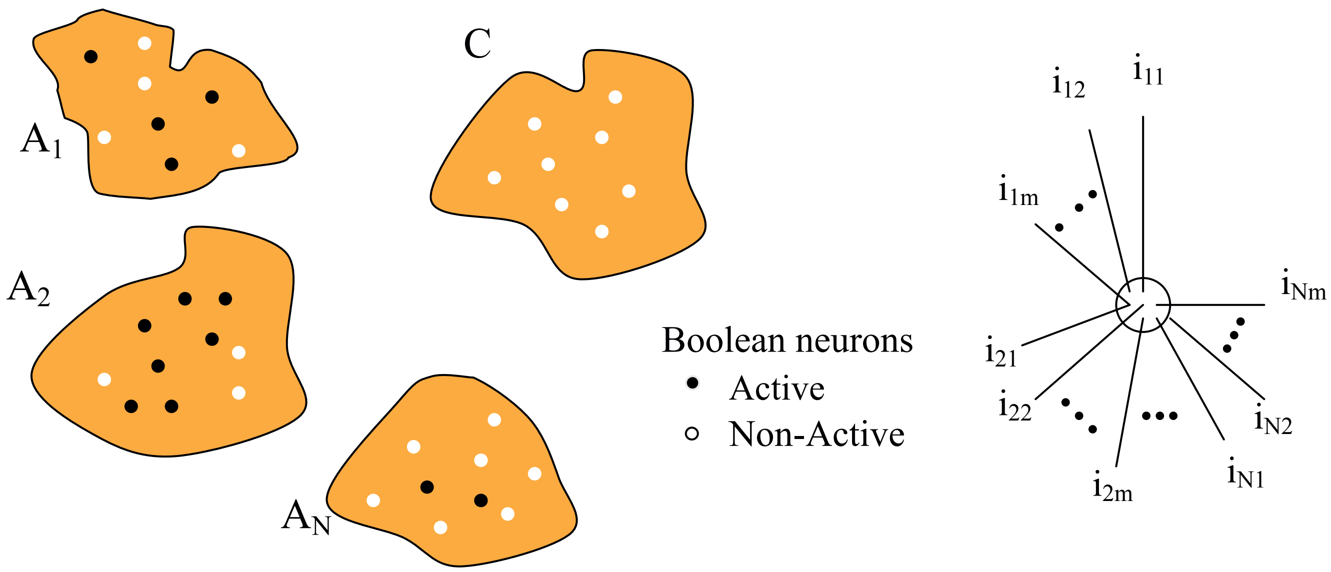



**Figure 1 – FBN Areas and Individual Neuron Structure: neurons are binary entities grouped into areas (or concepts); the value of each concept is given by the ratio of active neurons; each neuron has *mxN* inputs that can randomly sample the value of *m* neurons from each of *N* antecedent concepts**

Recent developments use the "non-taught" state of FF, and an additional Emotional Layer to deal with validation, and solve dilemmas and conflicting information [21].

*3.1 Inference*

Inference proceeds in the following way: each consequent neuron samples each of the antecedent areas using its *m* inputs. Note that *m* is always much smaller than the number of neurons per area. As we have seen in Section 3., for rules with *N* antecedents and a single consequent, each neuron has *N.m* inputs. FCR rules have a single antecedent, therefore, each consequent neuron will have *m* inputs and up to *m+1* internal memories. The single operation carried out by each neuron is the combinatorial count of the number of activated inputs from every antecedent (in the single antecedent case, this operation is reduced to counting the active inputs). Neurons have a unitary memory (FF) for each possible count combination, and its value will be compared with the corresponding sampled value. If the FF corresponding to the sampled value of all antecedents contains a '1', then the neuron output will be '1' (the neuron will be – or remain – activated); if the FF is '0', then the neuron output will be '0'. These operations can all be performed with classic Boolean AND/OR operations (any FBN can be implemented in hardware using basic logic gates). As a result of the inference process (which is parallel), each neuron will assume a binary value, and the inference result will be given by the neural activation ratio in the consequent area.

It has been proved [20] that, from these neuron micro operations, emerge a macro qualitative reasoning capability involving the concepts (fuzzy variables), which can be expressed as rules of type:

*IF Antecedent$_1$ is A$_1$ AND Antecedent$_2$ is A$_2$ AND …THEN Consequent is C$_i$,*

where *Antecedent$_1$*, *Antecedent$_2$,..*, *Antecedent$_2$* are fuzzy variables and *$A_1$, $A_2$; …, $C_i$* are linguistic terms with binomial membership functions (such as, "small", "high", etc.).

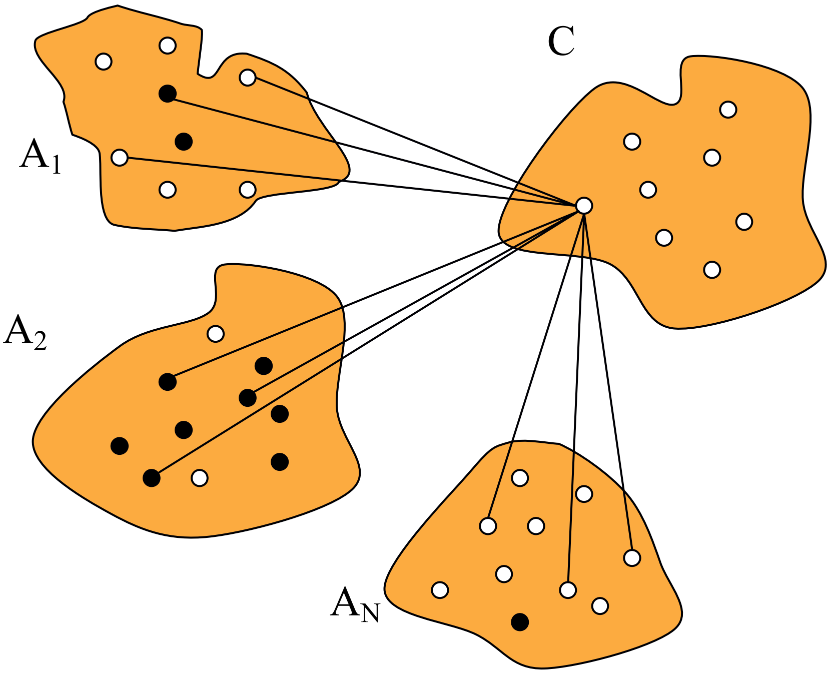



**Figure 2 – Inference of a single Neuron in a FBN: In this example, the neuron samples m=3 neurons in each of its N antecedent areas and counts how many of those neurons are active in each area. The combinatorial combination of the counts will be compared to a stored (previously learned) value that indicates if the neuron should or not be activated.**

This conclusion can be extracted from the interpretation of the consequent activation ratio, which is given by [20]:

$$\sum_{i1=0}^{m} .. \sum_{iN=0}^{m} \prod_{j=1}^{N} \binom{m}{ij} p_j^{\,ij} . (1-p_j)^{m-ij} . pr(i1,..,iN) \tag{7}$$

Where:

- $ij$ is the number of possible activated inputs to a single consequent neuron coming from antecedent $j$ (ranging from 0 to m);
- $p_j$ is the activation ratio of antecedent j, which, by definition varies between 0 and 1;
- pr($i1,\ldots,iN$) is the probability of FF associated with joint antecedent count ($i1,..,iN$) to be at "1" (the value of these FF are established during the learning phase and their activation probability depend on the experiments on that phase).

At the macroscopic/network level, this equation represents an emergent behaviour, which can be viewed as a fuzzy qualitative reasoning obtained from the microscopic neural operations (Boolean operations). In order to establish this fuzzy reasoning, the algebraic product and the bounded sum are interpreted respectively as the t-norm and t-conorm fuzzy operations. To this purpose the equation above should be interpreted as follows:

The activation ratios, $p_j$, are the input variables, and are fuzzified through binomial membership functions of the form

$$\binom{m}{ij} p_j^{ij}.(1-p_j)^{m-ij}. \qquad (8)$$

The evaluation of the expression for a given $p_j$ represents the membership degree of $p_j$ in that fuzzy set.

By definition of the above t-norm, the product of the terms, $\prod_{j=1}^{N}$, represents the fuzzy intersection of the *N* antecedents.

Considering the consequent fuzzy sets as normal singletons (amplitude "1") at the consequent UoD values, p(*i1*,..,*iN*), it follows that the equations represent the defuzzification by the Centre of Area method .

*3.2 Learning*

Learning is performed by exposing the net to experiments and modifying the internal binary memories of each consequent neuron according to the activation of the *m* inputs (per antecedent) and the state of that consequent neuron. Each experiment will set or reset the individual neuron's binary memories. Since FBN operation is based on random input samples for each neuron, learning (and inference) is a probabilistic process. For each experiment, a different input configuration (defined by the input areas specific samples) is presented to each and every of the consequent neurons, and addresses one and only one of the internal binary memories of each individual neuron. Updating of each binary memory value depends on its selection (or not) and on the logic value of the consequent neuron. This may be considered a Hebbian type of learning [11][25] if pre and post-synaptic activities are given by the activation ratios: if the concept of weight is made equivalent to the probabilities *p* of the internal flip-flops, then, when input-output correlation between a given input and a given output is positive, the corresponding weight increases as in Hebbian learning. Proof that the network converges to a taught rule, and a more detailed description of the learning process can be found in [22].

Although various learning strategies can be used, one considers the interesting case when non-selected binary memories maintain their state and selected binary memories take the value of consequent neuron. The corresponding updating equation is (indexes are not represented):

$$p(t+1)-p(t) = P^a . (p_{out} - p(t)) \qquad (9)$$

Where:

- p(t) is the probability of a given FF to be at "1" at time t;
- $P^a$ is the probability of activating the decoder output associated with p in the experiment;
- $p_{out}$ is the probability of one consequent neuron to be activated.

This case corresponds to a kind of Grossberg based learning [25] (which itself is a modified Hebbian learning), although more complex, since $P^a$ is variable with time (in the traditional Grossberg learning, $P^a$ is a constant [25]).

It has also been proved [18] that a FBN is capable of learning a set of different rules without cross-influence between different rules, and that the number of distinct rules that the system can effectively distinguish (in terms of different consequent terms) increases with the square root of the number m.

Finally, it has been proved that a FBN is a Universal Approximator [19], since it theoretically implements a Parzen Window estimator [13]. This means that these networks are capable of implementing any possible multi-input single-output function of the type: $[0,1]^n x[0,1]$.

These results give the theoretical background to establish the capability of these simple binary networks to perform qualitative reasoning and effective learning based on real experiments.

## 4. Qualitative Optimal Completion of Fuzzy Causal Rule Bases

The option to choose FBN to optimize raw FCR data was based on the fact that FBN properties as qualitative universal approximators could be used to allow a seamless and data independent optimization process, where rule learning and rule completion would be integrated in a single technique that is almost independent of the data source (single expert, multiple experts, or quantitative data). Moreover, we will provide examples which show that, with minor modifications, all sources can be used simultaneously.

In order to use FBN in the optimization process, the antecedent and consequent linguistic term set of the variables involved in the causal relation must be properly defined *a-priori*: even knowing that FBN have the capability of automatically extracting linguistic membership functions from raw quantitative data, these membership functions do not abide with the strict restrictions necessary in FCR. Therefore, one cannot use this capability to optimize the FCR. The centroid of each linguistic term membership function must also be made available.

The following sections detail the process of optimal FCRb completion using FBN.

### *4.1 Single Expert Knowledge FCRb Qualitative Optimal Completion*

Whenever FCR knowledge is obtained from a single expert, all provided rules are considered valid, unless the expert states its uncertainty regarding specific rules. If the expert does not provide a complete rule base then the rule base must be necessarily completed before it can be used (this is a very common situation due to the high number of linguistic terms usually involved in FCR). FBN mesh based structure gives them a good generalization capability. Even small sized FBN can automatically interpolate values from large areas where training data was missing. For example, a FBN with 128 neurons per area, each with 25 inputs, can properly cover $1/\sqrt{25}$=20% of the input area for each provided crisp input [18]. This is a theoretical limit, and in practice we can obtain even better coverage. Therefore we can use such FBN to complete any FCRb that can be described by 5 evenly spaced rules. It is important to note that, even if 5 rules are sufficient to describe the relation, the causal rule base still needs to be completed because 11 or 13 linguistic term sets are common (and necessary) in RB-FCM concepts (fuzzy variables) [1][5][6][7][8]. The procedure is as follows:

1. Use a FBN with one antecedent and one consequent area. Define 128 neurons per area with 25 inputs each and use maximum granularity. Although a larger FBN could provide a finer approximation degree, experiments show that these settings provide a good compromise between computer performance and results (see section 5);
2. Obtain the centroid of the antecedent and consequent linguistic terms ($x_i$ ,$z_k$) for each available expert rule;
3. Use all $x_i$ ,$z_k$ as training data for the FBN. Since all rules are considered valid, there is no need to use the FBN validation mechanisms and emotional layer. In such a FBN, experiments show that twenty training epochs are sufficient to produce stable results;
4. After training completion, the resulting FBN behaves as a qualitative approximator for the FCR in all UoD;

5. Too obtain the consequent of a missing rule ($C_j$), one has to feed the FBN with the centroid of the antecedent linguistic term of that rule. Since FBN are probabilistic, one should infer the FBN several times and average the results to obtain $z_j$. The chosen consequent, $C_j$, will be the one where $z_j$ has the highest membership degree;
6. A smooth qualitative interpolated completion is guaranteed as long as at least 5 evenly spaced rules were given by the expert [18], but even with 3 evenly spaced rules is possible to obtain good results (see section 5). Since FBN provide a way to verify the validity of a certain result (based on the ratio of taught/non taught neurons that were used to infer the result) it is always possible to know how satisfactory the completion is.

We can see that the overall procedure is similar to previous completion methods, replacing known techniques with FBN learning and inference. The FBN properties as qualitative universal approximators, and their high generalization capabilities, allows them to obtain a complete FCRb where the previously missing rules are qualitatively interpolated from the original sparse rulebase. There are several obvious advantages in using this approach, like the lack of parameterization or the possibility of evaluating the obtained results, which are important to target users lacking strong mathematical knowledge. Other advantages will be shown in the results section.

### *4.2 Multiple Expert Knowledge FCRb Qualitative Optimal Completion*

Multiple expert knowledge optimization differs from the single expert case in step 3 of section 4.1: since experts might provide conflicting or incongruent information, one must use the FBN validation mechanisms and emotional layer to minimize their influence. Therefore extra parameterization (like attributing expert credibility) is required in this step, but the overall approach remains the same.

### *4.3 Quantitative Data FCRb Qualitative Optimal Completion*

When all available raw FCR data results from crisp uncertain measurements and/or observations, then all data is used to train and optimize the FBN. The FBN will behave as a qualitative universal approximator, and the rules can be obtained according to step 5 of section 4.1. When using the proposed FBN settings, completion is guaranteed as long as gaps no larger than 20% are left uncovered [18](section 4.1), but, once again, it is possible to obtain valid complete rule bases even with gaps up to 80% (see section 5., example 2). This is obviously highly dependent on the relation we are modelling (for a gap that size, the relation must be rather linear), but it is a fact that other proposed methods cannot deal with these cases at all (see section 5). Once again, one must note that the FBN is capable of automatically providing a measure on how trustworthy is the completion result.

### *4.4 Multi-Source FCRb Qualitative Optimal Completion*

Since quantitative data and expert data are basically handled the same way, it is possible to use both simultaneously. However, since expert knowledge provides at most a single data pair per expert and per rule, while quantitative data can be in the magnitude of thousands, it is necessary to weight the data according to its source during the training process. A simple method to do it is to train the expert knowledge for a number of epochs proportional to the magnitude difference between quantitative and expert data. This process obviously involves some extra parameterization in the FBN training. It is, however, a process that can be easily automated. All the remaining optimization process is maintained as in section 4.1.

## 5. Results

In [19] it was proved that FBN are universal approximators. However, those are theoretical results, where FBN size is not limited by practical applications. Therefore we must see how a FBN behaves in this regard when dealing with computationally treatable parameters like those proposed as an example in section 4.

*Example 1* – In Figure 3 we present a 128 neurons (with 25 inputs each) FBN approximation for the non-linear function used in [16]:

$$f = 0.1 + 0.4\sin(2\pi x + 1) \tag{10}$$

We trained the FBN for 40 epochs with the same 6 examples, and obtained an average error of only 4%. The results show the high learning and generalization capabilities of an FBN when dealing with non-linear functions and sparse training data. As a comparison, a TPE system needs 25 regions and 100 different examples to obtain a similar result [16]. If we wanted to translate the FBN results to a fuzzy rule base, the obtained error would be irrelevant even with a very high domain partition.

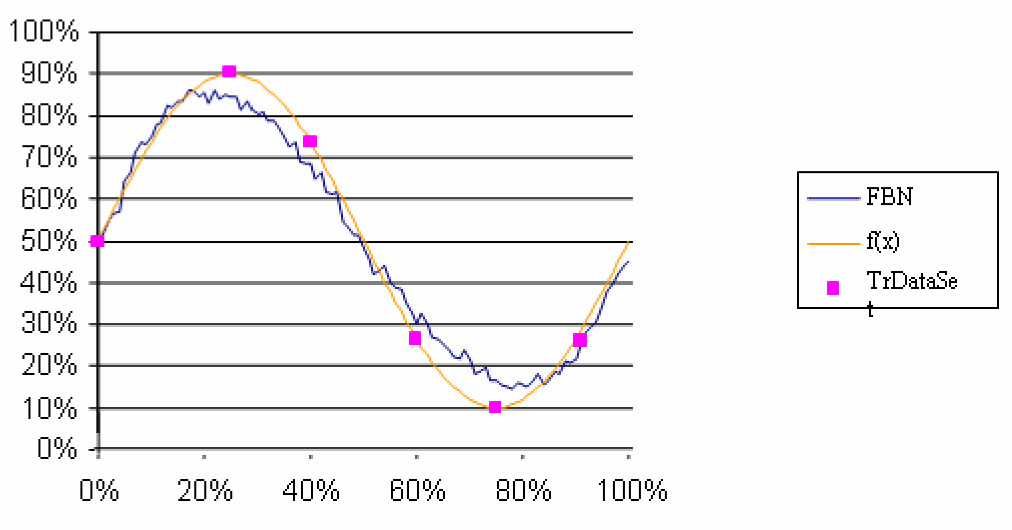



**Figure 3 – Medium sized FBN result as a function approximator with sparse training data**

*Example 2* – Regarding rule base completion, one of the major problems one has to cope when modelling a FCR, is the fact that experts will often only provide the minimum necessary information to describe the relation. For example, an economic expert expressing a qualitative causal relation between *Production* and *Price*, could probably simply state the following 2 rules[3]:

- *"If Production Increases, Then Price Decreases"*
- *"If Production Decreases, Then Price Increases"*

These rules are comprehensive enough for a human, and describe a simple offer/demand causal relation. There is no need for the expert to provide more rules since the additional information can be easily generalized by a human. Our problem is how to do it optimally and automatically. Consider a case where eleven different linguistic terms are defined in the fuzzy variable *Production* = {*Decreases_Very_Much*, *Decreases_Much*, *Decreases*, *Decreases_Few*, *Decreases_Very_Few*, *Maintains*, *Increases_Very_Few* *Increases_Few*, *Increases*, *Increases_Much*, *Increases_Very_Much*}.

[3] Note that this is just a simple example, and does not necessarily expresses a valid real world relation

Given that the relation is semantically linear and symmetric, this is obviously a pretty simple task for a human, and it should not be difficult to automate the procedure as long as the number or linguistic terms in *Price* is similar. Let us consider the simplest case, where linguistic terms of *Price* and *Oil_Production* are exactly the same (Figure 4). The intended result would certainly be the one expressed in the second column of Table 1.

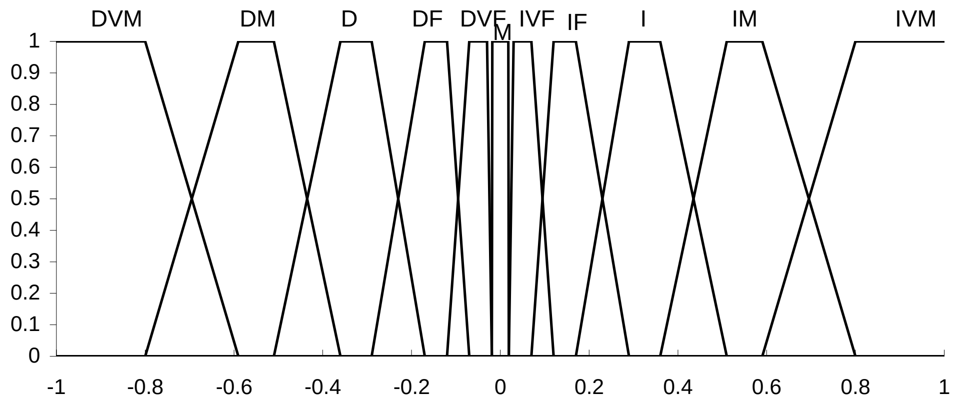



**Figure** 4 **– Membership functions for *Price* and *Production* linguistic terms**

However, even in such a simple example, the use of the completion techniques presented in section 2 will produce substantially different and unsatisfactory results due to the high sparseness degree of the available data. Table 1 presents the obtained results and compares them with the FBN approach we propose.

**Table 1 - Completion of a highly sparse FCRb using different approaches: Boldface values represent the initially available data; Shaded cells represent optimal completion.**

| | Optimal Solution | Region Growing | | Weighted Averaging | | FBN | | FBN(with outer rule extrapolation) | |
|---|---|---|---|---|---|---|---|---|---|
| ***Production*** | ***Price*** | ***Price*** | | ***Price*** | | ***Price*** | | ***Price*** | |
| -0.8 (DVM) | 0.8 (IVM) | 0.35 | I | 0.11 | IF | 0.32 | I | 0.78 | **IVM** |
| -0.55 (DM) | 0.55 (IM) | 0.35 | I | 0.08 | IVF | 0.32 | I | 0.54 | IM |
| **-0.35 (D)** | **0.35 (I)** | **0.35** | **I** | **0.35** | **I** | 0.30 | **I** | 0.40 | I |
| -0.15 (DF) | 0.15 (IF) | 0.35 | I | 0.03 | M | 0.12 | IF | 0.12 | IF |
| -0.08 (DVF) | 0.08 (IVF) | 0.35 | I | 0.02 | M | 0.08 | IVF | 0.08 | IVF |
| 0 (M) | 0 (M) | 0 | M | 0 | M | -0.04 | M | 0.01 | M |
| 0.08 (IVF) | -0.08 (DVF) | -0.35 | D | -0.02 | M | -0.09 | DVF | -0.09 | DVF |
| 0.15 (IF) | -0.15 (DF) | -0.35 | D | -0.03 | M | -0.15 | DF | -0.20 | DF |
| **0.35 (I)** | **-0.35 (D)** | -0.35 | **D** | **-0.35** | **D** | -0.30 | **D** | -0.36 | D |
| 0.55 (IM) | -0.55 (DM) | -0.35 | D | -0.08 | DVF | -0.38 | D | -0.54 | DM |
| 0.8 (IVM) | -0.8 (DVM) | -0.35 | D | -0.11 | DF | -0.38 | D | -0.76 | **DVM** |

With the Region Growing technique [16], each new rule is based solely on the closest neighbours. This is an iterative process where a rule consequent value is calculated by the average of its non empty neighbour consequent values. Therefore, due to the sparseness of available data, most rules will maintain the consequent values of that data, and we can see that the results are far from ideal.

Weighted Averaging [16] results show that this is a disastrous technique in such sparse rule bases. In this method all consequents are calculated simultaneously. Since only two rules are available, and each rule consequent is based on an average of the existing rules consequents weighted by the similarity (essentially based on the distance) of those rule antecedents, each rule basically cancels the other, and all obtained consequents (except for available data) either represent small variations or the absence of variation.

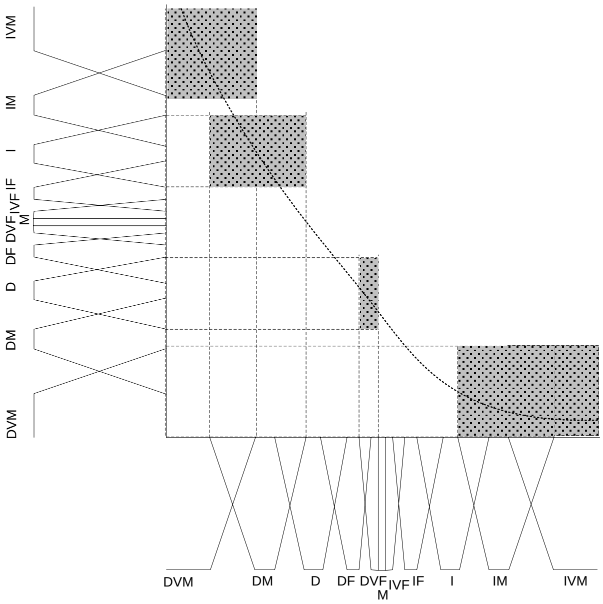
FBN optimization produces the best results, optimal in the centre regions, but far from ideal in the extreme regions of the UoD. This is due to the fact that those regions are not reached by the 20% coverage area of each training example. The immediate solution to avoid this problem is to always extrapolate this kind of knowledge (which forcefully represents a linear relation) to the outer rules by always replacing the linguistic terms provided by the expert for the outer antecedent linguistic terms. Last column of Table 1 shows that FBN will provide the optimal completion using this approach (the completion is optimal since it provides the same results as the original function); even if the interval without training data is close to 80% (the other 2 methods will still give inaccurate results).

The results show that previously presented completion techniques should not be used when data is too sparse, which often is the case in FCR optimization. The exception lies in the Interpolation by Similarity, with which we could obtain optimal results. However this would imply parameter tailoring for each particular case, therefore not complying with our main goals.

The previous example represents an extreme, although common, modelling FCR situation where the proposed approach can make a difference regarding optimization of raw data bases. However, one can find examples where FBN optimization produces good results where others can be rather unsatisfying even in more mainstream FCR optimization problems involving lower sparseness degrees. The problem with other methods in those situations is that they are still highly dependent on the location of the missing data [16]. In some simple cases, some of those methods present flaws, while the others behave well, being difficult to automatically select which one to use. Since our approach always relies on FBN properties as qualitative universal approximators, it is immune to these situations, even with nonlinear and non-symmetric relations or with cases where the number or syntax of consequent linguistic terms is not similar to those of the antecedent. The following example presents a rather mainstream case where our approach provides the optimal solution and the other methods fail to provide a proper solution.

*Example 3* – A non-linear and non-symmetric function was drawn based on the linguistic term set depicted in Figure **4**. The example case uses the same linguistic term set. A set of 5 rules was extracted directly from the graphic. These rules and the complete set of 11 linguistic terms were given to 6 human testers who were asked to complete the rule base. Figure **5** shows a pictorial representation for the proposed case.
All experts completed the FCRb with the same rules, except for input *Increase_Very_Few* (IVF), where opinions divided among *Decrease* and *Decrease_Much*.

Table 2 shows the results obtained by applying several techniques: Column 1 shows all antecedent linguistic terms; Boldface values represent supplied data; Column 2 shows the expert completion, and is our best shot for an optimal solution, since FCRbases reflect expert opinions; Columns 3, 4 and 5 show FCRb completion using several techniques.




**FCRb:**
*If X is DVM, Then Y is IVM*
*If X is DM, Then Y is I*
*If X is DVF, Then Y is D*
*If X is IM, Then Y is DVM*
*If X is IVM, Then Y is DVM*

**Figure 5 – 5 rule FCRb and linguistic term set provided to the experts (the function was omitted)**

**Table 2 - Completion of a FCRb using different approaches: Boldface values represent the initially available data; Shaded cells represent optimal completion. The only method to provide a result consistent with expert opinion is completion by FBN optimization.**

| *X* | Optimal (expert opinion) *Y* | Region Growing *Y* | | Weighted Averaging *Y* | | FBN(128/8inp) *Y* | |
|---|---|---|---|---|---|---|---|
| **-0.80 (DVM)** | **0.80 (IVM)** | **0.80** | **IVM** | **0.80** | **IVM** | 0.82 | IVM |
| **-0.55 (DM)** | **0.35 (I)** | **0.35** | **I** | **0.35** | **I** | 0.42 | I |
| -0.35 (D) | 0.08 (IF) | 0.35 | I | -0.04 | DVF | 0.1 | IF |
| -0.15 (DF) | -0.15 (DF) | 0 | M | -0.13 | DF | -0.24 | DF |
| **-0.08 (DVF)** | **-0.35 (D)** | **-0.35** | **D** | **-0.35** | **D** | -0.28 | D |
| 0 (M) | -0.35 (D) | -0.35 | D | -0.18 | DF | -0.34 | D |
| 0.08 (IVF) | -0.35/-0.55 (D/DM) | -0.35 | D | -0.21 | DF | -0.4 | D |
| 0.15 (IF) | -0.55 (DM) | -0.80 | DVM | -0.24 | DF | -0.52 | DM |
| 0.35 (I) | -0.80 (DVM) | -0.80 | DVM | -0.32 | D | -0.74 | DVM |
| **0.55 (IM)** | **-0.80 (DVM)** | **-0.80** | **DVM** | **-0.80** | **DVM** | -0.84 | DVM |
| **0.80 (IVM)** | **-0.80 (DVM)** | **-0.80** | **DVM** | **-0.80** | **DVM** | -0.90 | DVM |

The results expressed in

Table 2 and in Figure **6,** Figure **7** and Figure **8**, show that the FBN approach provides an optimal completion (since results are the same as those chosen by experts), while Region Growing differs from expert opinion in 3 out of 6 rules, and Weighted Averaging differs in 5 out of 6. Note the importance of these non-optimal completions: minor rule base differences in RB-FCM modelling can cause huge differences in scenario simulations and results due to the feedback cycles.

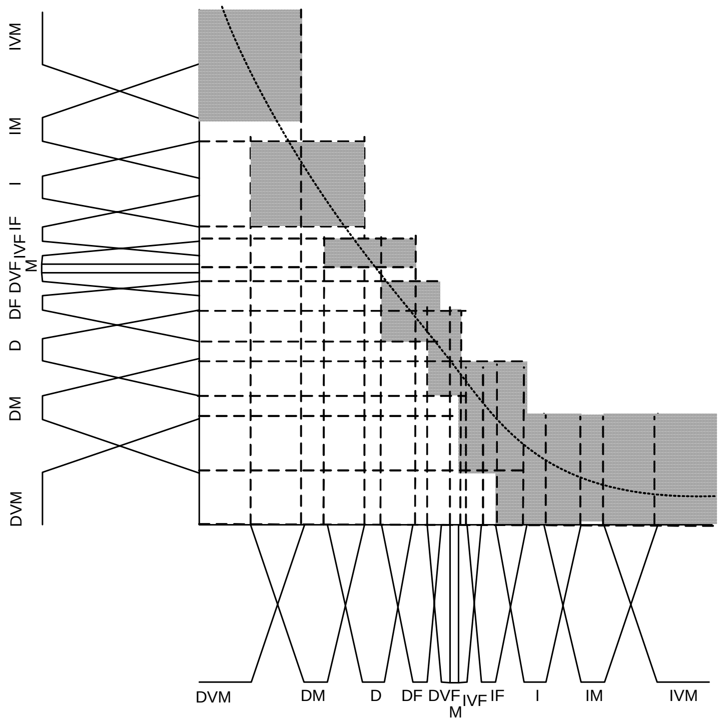



**Figure 6 – FCRb completion by FBN**

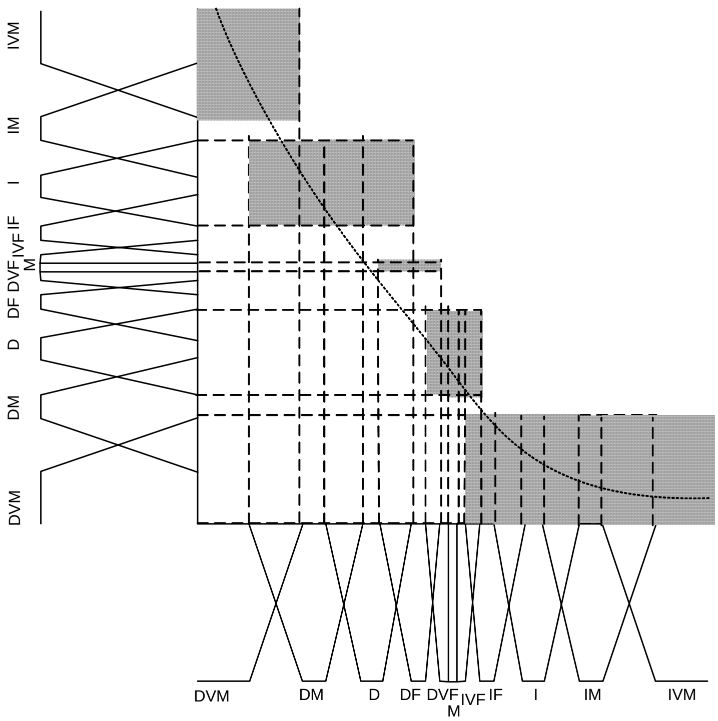



**Figure 7 – FCRb completion using Region Growing**

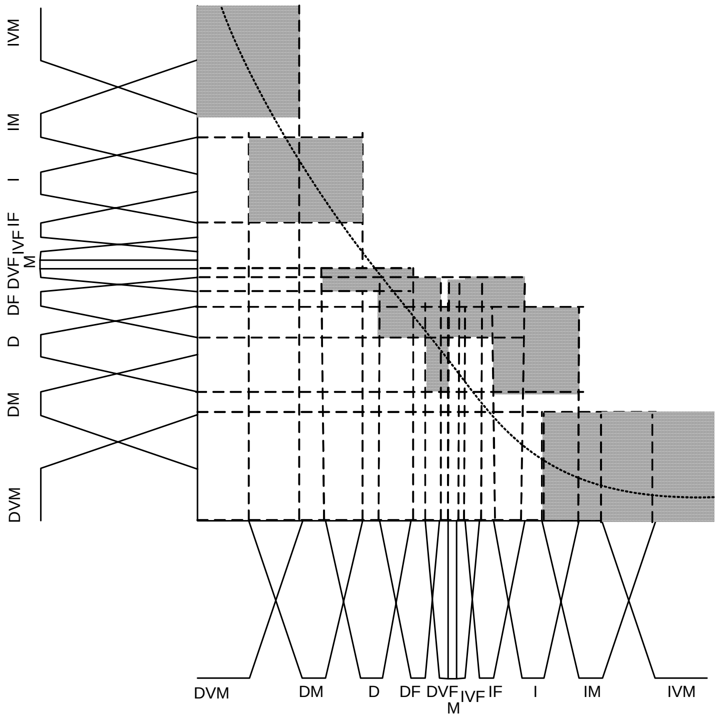



**Figure 8 – FCRb completion using Weighting Averaging**

*Example 4* – This example shows FBN optimization of a FCRb when quantitative source data is used obtain the rule base. The FBN was trained using 30 randomly scattered data points concentrated around the 5 areas corresponding to 5 rules of the previous example. Figure 9 shows the training data set and the FBN result. FCRb completion using these results was also optimal (the same as expert panel completion).

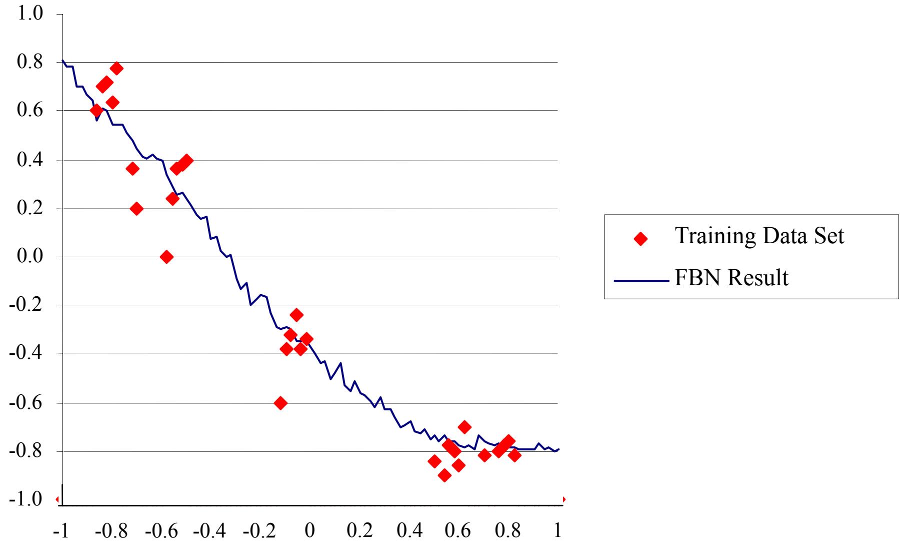



**Figure 9 – FCRb Optimal Completion by application of FBN on quantitative source data**

In Figure 9 one can observe that small dimension FBNs, like the ones we use, do not provide "smooth" approximations. The local variations are usually not critical, since they are much smaller than the support of the consequent linguistic terms. However, since one uses a single point of the antecedent linguistic term (the centroid) to compute the consequent of an unavailable rule, it is not impossible to obtain a less correct completion if the centroid coincides with one of the local maxima. These cases can be minimized by increasing the number of times one infers the consequents in step 5 of the optimization process. Other possible solution is proposed in section 7. One must note however, that these cases should be quite rare (none has ever been detected in our experiments).

*Example 5* – Finally we present an example where quantitative data and expert knowledge are used simultaneously as source data. In this example, the experts provided the 2 outermost rules (the extreme cases are usually the ones where expert have less difficulties to express their knowledge), and the rest of the training data consisted in quantitative samples similar to those presented in Example 3 (20 data points were used). Training used the approach described in section 4.3. The results are shown in Figure 10. Once again, FCRb completion using these results was optimal (the same as expert panel completion), and quite similar to the previous example (differences exist and are explained by differences in training data, but completion result is identical).

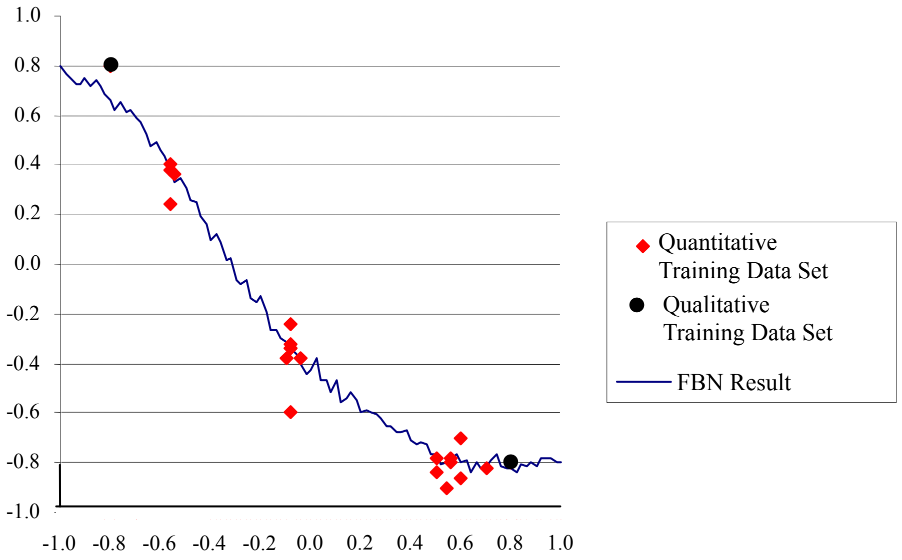



**Figure 10 – FCRb Optimal Completion by application of FBN on Multi-source Data (qualitative and quantitative)**

## 6. Generalization

Even though we focused on the optimization and completion of Fuzzy Causal rule bases, the methods we presented in this paper can be generalized to all fuzzy rule bases. However, when we generalize the process, we loose performance due to the exponential increase of FBN size when the number of antecedents increases, even if this loss can be minimized by using granularity in FBN's internal memories.

Our method still has advantages when dealing with uncertain and sparse quantitative data since FBN are still capable of extracting qualitative rules and of creating and validating complete rule bases, with the additional advantage of not needing an *a priori* linguistic term definition (except when, as with FCR, the linguistic terms must adhere to some restrictions).

## 7. Conclusions and Future Developments

We have presented a FCRb optimization method based on FBN. This method allows a seamless and data source independent optimization process, with good generalization capabilities, where rule learning and rule completion are integrated in a single technique. In order to minimize eventual completion ill-effects caused by local oscillations in the FBN result, future developments include granularization tuning and the use of the entire support set of antecedent linguistic terms in consequent inference (as opposed to using only their centroid).

## 8. References


[1] Carvalho, J.P., Tomé, J.A., "Fuzzy Mechanisms For Causal Relations", Proceedings of the Eighth International Fuzzy Systems Association World Congress, IFSA'99, Taiwan, 1999, 1009-1013.

[2] Carvalho, J.P., Tomé, J.A., "Interpolated Linguistic Terms: Uncertainty Representation in Rule Based Fuzzy Systems", Proceedings of the 22nd International Conference of the North American Fuzzy Information Processing Society, NAFIPS2003, Chicago, 2003, 93-98.

[3] Carvalho, J.P., Tomé, J.A., "Interpolated Linguistic Terms", Proceedings of the 23rd International Conference of the North American Fuzzy Information Processing Society, NAFIPS2004, Banff, Canada, 2004, 151-156.

[4] Carvalho, J.P., Tomé, J.A., "Issues on the Stability of Fuzzy Cognitive Maps and Rule-Based Fuzzy Cognitive Maps", Proceedings of the 21st International Conference of the North American Fuzzy Information Processing Society, NAFIPS2002, New Orleans, 105-110.

[5] Carvalho, J.P., Tomé, J.A., "Mapas Cognitivos Baseados em Regras Difusas: Modelação e Simulação da Dinâmica de Sistemas Qualitativos", PhD thesis, Instituto Superior Técnico, Technical University of Lisbon, Portugal, 2002, 97-156.

[6] Carvalho, J.P., Tomé, J.A., "Qualitative Modelling of an Economic System using Rule Based Fuzzy Cognitive Maps", FUZZ-IEEE 2004 - IEEE International Conference on Fuzzy Systems, Budapest, 2004, 659-663.

[7] Carvalho, J.P., Tomé, J.A., "Rule Based Fuzzy Cognitive Maps - Fuzzy Causal Relations", Computational Intelligence for Modelling, Control and Automation – Evolutionary Computation & Fuzzy Logic for Intelligent Control, Knowledge Acquisition & Information Retrieval, Edited by M. Mohammadian, IOS Press ISBN 90 5199 4745, 1999, 276-281.

[8] Carvalho, J.P., Tomé, J.A., "Rule Based Fuzzy Cognitive Maps – Qualitative Systems Dynamics", Proceedings of the 19th International Conference of the North American Fuzzy Information Processing Society, NAFIPS2000, Atlanta, 2000, 407-411.

[9] Cross,V., Sudkamp,T., "Sparse data and rule base completion", Proceedings of the 22nd International Conference of the North American Fuzzy Information Processing Society, NAFIPS2003, Chicago, 2003, 81-86.

[10] Dubois, D., Prade, H., "On Fuzzy Interpolative Reasoning", International Journal of General Systems, vol. 28, 1999, 103-114.

[11] Hebb, D., "The Organization of Behaviour: A Neuropsychological Theory", John Wiley & Sons, 1949

[12] Koczy, L., Hirota, K., "Approximate Reasoning by Linear Rule Interpolation and General Approximation", International Journal of Approximate Reasoning, Vol 9 (3), 1993, 197-225.

[13] Parzen, E., "On Estimation of a probability density function and mode" Ann. Math. Stat., 33, 1962, 1065-1072.

[14] Qiao, W.Z, Masaharu, M., Yan, S., "An improvement to Kóczy and Hirota's interpolative reasoning in sparse fuzzy rule bases", International Journal of Approximate Reasoning, Vol 15 (3), 1996, 185-201.

[15] Sudkamp,T., "Similarity, Interpolation and Fuzzy Rule Construction", Fuzzy Sets and Systems, vol.58, 1993, 73-86.

[16] Sudkamp,T., Hammell,R.J., "Interpolation, Completion, and Learning Fuzzy Rules", IEEE Transactions on Systems, Man, and Cybernetics, Vol. 24, 2, 1994, 332-342.

[17] Taber, R., Yager, R., Helgason, C., "Small Sample Effects on the Equilibrium Behaviour of Combined Fuzzy Cognitive Maps", Proceedings of the FUZZ-IEEE2001, Melbourne, Australia, 2001, 1567-1572.

[18] Tomé, J.A. and Carvalho, J.P. "Rule Capacity in Fuzzy Boolean Networks", Proceedings of the 21st International Conference of the North American Fuzzy Information Processing Society, NAFIPS2002, New Orleans, 2002, 124-128.

[19] Tomé, J.A., "Counting Boolean Networks are Universal Approximators" Proceedings of the 1998 Conference of NAFIPS, Florida, 1988, 212-216.

[20] Tomé, J.A., "Neural Activation ratio based Fuzzy Reasoning", Proceedings of the IEEE World Congress on Computational Intelligence, Anchorage, 1988, 1217-1222.

[21] Tomé, J.A., Carvalho, J.P., "Decision Validation and Emotional Layers on Fuzzy Boolean Networks", Proceedings of the 23rd International Conference of the North American Fuzzy Information Processing Society, NAFIPS2004, Banff, Canada, 2004, 136-139.

[22] Tomé, J.A., Tomé, R,. Carvalho, J.P., "Extracting Qualitative Rules from Observations – A Practical Behavioural Application", WSEAS Trans. On Systems, Issue 8, Vol 3., 2004, 2721-2726.

[23] Ughetto, L., Dubois, D., Prade, H., "Fuzzy Interpolation by Convex Completion of Sparse Rule Bases", Proc. Of the Ninth IEEE International Conference on Fuzzy Systems, San António, TX, USA, 2000, Vol 1., 465-470.

[24] Wang, L.X., Mendel, J.M., "Generating fuzzy rules by learning from examples", IEEE Trans. System, Man and Cybernetics, vol. 22, no. 6, 1992, 1414-1427.

[25] Lin, C.-T. and Lee, C.S.G., "Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems", Prentice-Hall, 1996, 301-309.