Test- and Rating Strategies for Data Based Rule Generation*

Holger Jessen and Timo Slawinski

June 1998

Abstract

The paper presents new strategies for testing and rating the relevance of rules in the Fuzzy–ROSA (Rule Oriented Statistic Analysis) method for data based rule generation. Specific characteristics and differences between the proposed strategies are pointed out.

Keywords: Fuzzy systems, rule based modeling, data based modeling, relevance test and rating, Fuzzy–ROSA method

1 Introduction

Modeling of a given process can be carried out either theoretically or empirically. The theoretical approach is based on a theory and existing knowledge about the process. The empirical approach uses measured input/output data.

The Fuzzy–ROSA (Rule Oriented Statistic Analysis) method is an empirical approach using fuzzy–if–then–rules to describe the observed behaviour of a process [1, 2]. The if–then–rules have the form

$$\text{IF} \quad p_k \quad \text{THEN} \quad c_k \tag{1}$$

with k indicating the k–th rule. The premise part p_k of the rule is a statement on the input vector $\underline{x}$ and the conclusion part c_k is a statement on the output y of the forms

$$p_k = \{(x_i = a_{i,l}) \wedge \quad \cdots \quad \wedge (x_j = a_{j,m})\} \qquad \text{and} \qquad c_k = \{y = b_z\} \tag{2}$$

*This research was sponsored by the Deutsche Forschungsgemeinschaft (DFG), as part of the Collaborative Research Center 'Computational Intelligence' (531) of the University of Dortmund

where x_i is the i-th component of $\underline{x}$ and $a_{i,l}$ is the l-th corresponding linguistic value for this component. Accordingly, b_z is the z–th linguistic output value. The linguistic values a and b are defined by triangular or trapezoidal membership functions (fuzzy–sets).

To generate a fuzzy model from the given input/output data means to find appropriate rules. The Fuzzy–ROSA method is based on the idea to evaluate the correctness and relevance of every potential rule (hypothesis) by a statistical test. If the hypothesis proves to be relevant and correct with respect to the given data, it is accepted as a rule and added to the rule base. An index rating the relevance of the rule according to the test is attached to each rule [3, 4].

The number of possible hypotheses depends on the number of input variables (components of $\underline{x}$) and the number of linguistic values per variable and can be very large because of the combinatorial complexity. To receive a small rule base not only complete rules considering all components of the input vector in their premise part but also generalizing rules are included in the rule generation process. Generalizing rules have less complex premise parts by neglecting components of the input vector and thus cover a wider range of the observation data. This also increases the transparency of the rule base. The generated rule base usually will be further reduced and simplified by several rule reduction concepts. These concepts strongly depend on the rating of the rules calculated during the generation process. Rules with a lower rating are considered to be less relevant and correct according to the observed data and therefore are more likely to be reduced [5].

This paper proposes new strategies for the test and rating of potential rules. In the following section after a short introduction the different strategies are presented in mathematical detail. The original relevance index for the Fuzzy–ROSA method is compared to the new strategies. The last section gives possible applications and points at further research conducted by the authors.

2 Rule Test and Rating Strategies

From the Fuzzy–ROSA method point of view the quality of a rule base consists in the quality of its individual rules. The quality of a model, on the other hand, is determined by the modeling objective. Different modeling strategies will be useful depending on whether the model is to be explanatory, descriptive or predictive.

Setting up a good rule based model therefore requires efficient strategies for generating rules of high quality. The new rule test and rating strategies presented in the following are intended to serve as tools for a goal–oriented approach to generate fuzzy rule based models.

2.1 Relevance Index (RI)

The Relevance Index was originally developed for the Fuzzy–ROSA method for modeling human behaviour [6]. According to this index, a rule is relevant, if the constrained probability $P(c|p)$ of its conclusion part c given the premise part p exceeds the unconstrained probability $P(c)$ in the given data.

The probability of the fuzzy conclusion part c is estimated by:

$$\hat{P}(c) = \frac{\sum_n \mu(y_n = b_z)}{N} \tag{3}$$

where y_n is the n–th output value in the observed data, b_z is the linguistic value in the conclusion part of the rule and N is the number of observation data and $\mu(y_n = b_z)$ is the truth value or degree of membership of y_n being b_z. The constrained probability $P(c|p)$ is estimated by

$$\hat{P}(c|p) = \frac{\sum_n \mu(x_{n,i} = a_{i,l}) \cdot \quad \cdots \quad \cdot \mu(x_{n,j} = a_{j,m}) \cdot \mu(y_n = b_c)}{\sum_n \mu(x_{n,i} = a_{i,l}) \cdot \quad \cdots \quad \cdot \mu(x_{n,j} = a_{j,m})} \tag{4}$$

Here, $x_{n,i}$, $x_{n,j}$ are the i-th and j-th component of the n-th input vector $\underline{x}_n$, $\mu(x_{n,i} = a_{i,l})$ is the truth value of $x_{n,i}$ being $a_{i,l}$. In this expression, the logical $\wedge$ in equation (2) is implemented by the algebraic product.

To further improve the estimate of these probabilities, the confidence intervals for the probabilities are calculated for a given confidence level α and used instead of the estimated probabilities [7]. $\hat{P}(c)$ is replaced by the upper bound, $\hat{P}(c|p)$ by the lower bound of its confidence interval, respectively. A rule is accepted, if

$$V_l(c|p) > V_u(c) \tag{5}$$

with $V_l(c|p)$ and $V_u(c)$ the lower and upper bound of the according confidence intervals for the probabilities.

If

$$V_l(c) > V_u(c|p) \tag{6}$$

the conclusion part is inverted and a negative rule

$$\text{IF} \quad p_k \quad \text{THEN} \quad c_k \quad \text{FORBIDDEN} \tag{7}$$

is added to the rule base [2, 8].

The index rating the relevance of the positive or negative rule respectively is calculated as

$$J_{RI} = \left\{ \begin{array}{ll} \dfrac{V_l(c|p) - V_u(c)}{1 - \hat{P}(c)} & \text{if} \quad V_l(c|p) > V_u(c) \\ \dfrac{V_l(c) - V_u(c|p)}{\hat{P}(c)} & \text{if} \quad V_l(c) > V_u(c|p) \end{array} \right\} \tag{8}$$

In equation (8) the distance of the confidence interval bounds is normed by its theoretical maximum calculated from the estimated probability $\hat{P}(c)$.

Confidence intervals contain the true value of a population with a $1 - \alpha$ confidence. Using the lower or upper bound of the confidence interval for each probability, therefore produces a more reliable estimate for the relevance of a rule. Also, a confidence interval supported by few data is larger than one for many data so that frequent situations in the observation data are more likely to produce relevant rules. The value of α can be adjusted to take into account the amount of observation data.

2.2 Normalized Hit Rate

In some applications, especially when there are few observed data available, the normalized hit rate, which is simply the estimate of the constrained probability, is more useful than the relevance index. The rule test and rating according to the normalized hit rate is independent of the probability of the conclusion part in the observed data and does not take the size of the database into account. The test is defined using equation (4):

$$\begin{array}{lllllll} \hat{P}(c|p) > \Theta & \rightarrow & \text{IF} & p & \text{THEN} & c & \\ \hat{P}(c|p) < \Theta & \rightarrow & \text{IF} & p & \text{THEN} & c & \text{FORBIDDEN} \end{array} \tag{9}$$

A positive rule is accepted, if the probability exceeds a choosable treshold Θ. $\Theta = 0.5$ means that a rule is accepted, when it is more likely to be true than false. An accepted rule is rated by mapping its probability to the interval $[0, 1]$:

$$J_{NHR} = \left\{ \begin{array}{ll} \frac{\hat{P}(c|p) - \Theta}{1 - \Theta} & \text{if} \quad \hat{P}(c|p) > \Theta \\ \frac{\Theta - \hat{P}(c|p)}{\Theta} & \text{if} \quad \hat{P}(c|p) < \Theta \end{array} \right\} \tag{10}$$

2.3 Confident Normalized Hit Rate

When sufficient observation data are available, it might be desirable to include the support of a rule from the given database in the rule test and rating strategy. This will be the case when relevant rules which are supported by many observation data are to be separated from random effects which occur only once or twice in the data. A reasonable test and rating strategy is to replace the probability in equation 10 by its upper and lower confidence interval bounds respectively. This combines the concepts of hit rate and relevance. Accepted rules are more likely to be true than false with a $1 - \alpha$ confidence. The rule test is defined as

$$\begin{array}{lllllll} V_l(c|p) > \Theta & \rightarrow & \text{IF} & p & \text{THEN} & c & \\ V_u(c|p) < \Theta & \rightarrow & \text{IF} & p & \text{THEN} & c & \text{FORBIDDEN} \end{array} \tag{11}$$

The rule is rated as:

$$J_{CNR} = \left\{ \begin{array}{ll} \frac{V_l(c|p) - \Theta}{1 - \Theta} & \text{if} \quad V_l(c|p) > \Theta \\ \frac{\Theta - V_u(c|p)}{\Theta} & \text{if} \quad V_u(c|p) < \Theta \end{array} \right\} \tag{12}$$

2.4 Relevant Hit Rate

The relevant hit rate is a combination of the relevance index test and the normalized hit rate rating strategy. The test is defined by equations (5) and (6) respectively:

$$\begin{array}{lllllll} V_l(c|p) > V_u(c) & \rightarrow & \text{IF} & p & \text{THEN} & c & \\ V_u(c|p) < V_l(c) & \rightarrow & \text{IF} & p & \text{THEN} & c & \text{FORBIDDEN} \end{array} \tag{13}$$

The rating is:

$$J_{RHR} = \left\{ \begin{array}{ll} \hat{P}(c|p) & \text{if} \quad \hat{P}(c|p) > \hat{P}(c) \\ 1 - \hat{P}(c|p) & \text{if} \quad \hat{P}(c|p) < \hat{P}(c) \end{array} \right\} \tag{14}$$

2.5 t–Test

Often the output variable of the observed data is continuous on an interval scale and the mean of the output value given an input situation is the desired information.

The t–test is a test strategy on hypotheses about means [9]. In the t–test the mean $\bar{y}_p$ of the output value y_p given the premise part p of a fuzzy–if–then rule is calculated and tested rather than the probabilities of the premise or conclusion part of a fuzzy rule.

Essentially the t–test is a test, whether the mean $\bar{y}_p$ of those output values for which the premise part is true is significantly different from the mean $\bar{y}$ of all output values. If $\bar{y}_p$ and $\bar{y}$ are significantly different, the rule is considered to be relevant. Figure 1 illustrates the basic idea of the test.

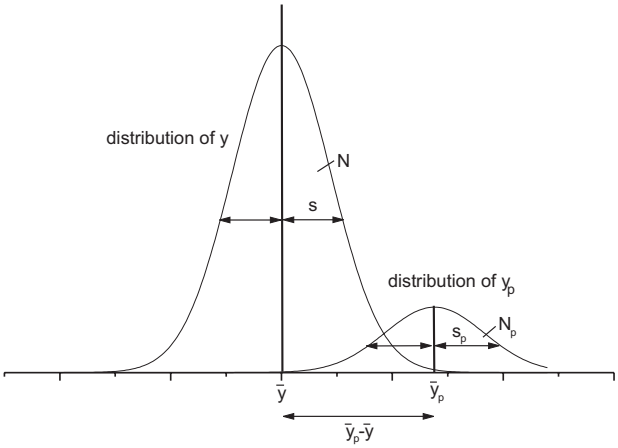

Figure 1: t–test

The t–test therefore concentrates on the output variable y rather than the constrained probability $P(c|p)$ of a fuzzy rule.

According to the t–test, a rule is accepted, when

$$\left| \frac{\bar{y}_p - \bar{y}}{\sigma_{\bar{y}}} \right| > t(N_p, \alpha) \tag{15}$$

with $\bar{y}_p$ the weighted mean of the output data given the premise p:

$$\bar{y}_p = \frac{\sum_n \mu(x_{n,i} = a_{i,l}) \cdot \quad \cdots \quad \cdot \mu(x_{n,j} = a_{j,m}) \cdot y_n}{\sum_n \mu(x_{n,i} = a_{i,l}) \cdot \quad \cdots \quad \cdot \mu(x_{n,j} = a_{j,m})} \tag{16}$$

and N_p the number of output data given the premise p:

$$N_p = \sum_n \mu(x_{n,i} = a_{i,l}) \cdot \quad \cdots \quad \cdot \mu(x_{n,j} = a_{j,m}) \tag{17}$$

$\bar{y}$ is the mean of the N output data:

$$\bar{y} = \frac{\sum_n y_n}{N} \tag{18}$$

The standard deviation $\sigma_{\bar{y}}$ of the distribution of $\bar{y}$ is estimated by:

$$\sigma_{\bar{y}} = \frac{s}{\sqrt{N_p}} \tag{19}$$

where s is the standard deviation of the output data

$$s = \sqrt{\frac{\sum_n (y_n - \bar{y})^2}{N-1}} \tag{20}$$

$t(N_p, \alpha)$ is the critical value to reject the null hypothesis at the confidence level α. It can be calculated using approximations or taken from a table of t–distributions.

A possible rating of an accepted rule is to consider the constrained standard deviation s_p of the output value given the input situation p:

$$J_t = e^{-\, s_p/s} \tag{21}$$

The e-function is used to map the relative standard deviation s_p/s to $[0, 1]$. s_p is calculated as

$$s_p = \sqrt{\frac{\sum_n \mu(x_{n,i} = a_{i,l}) \cdot \quad \cdots \quad \cdot \mu(x_{n,j} = a_{j,m}) \cdot (y_n - \bar{y}_p)^2}{(N_p - 1)}} \tag{22}$$

The conclusion part c is chosen as b_z, so that $\mu(\bar{y} = b_z) \rightarrow \max$.

3 Concluding Remarks

Three new rule test and rating strategies are presented in this paper.

The normalized hit rate is a test and rating strategy which tests the validity of fuzzy–if–then-rules irrespective of the distribution of the output data. The number of data supporting each rule is not considered for testing and rating so that rules are generated

even when their database is extremely small. This may lead to an overfitting of the model to rare input situations.
The normalized hit rate is appropriate for generating as many valid rules as possible, especially when there are vew observed data available. It appears to be most feasible for key–field modeling.

The confident normalized hit rate is more conservative than the normalized hit rate concerning the database of a rule. By using the bounds of the confidence intervals instead of the probabilities the generated rule base is not only valid but also relevant. Relevance in this sense means that the generated rules were observed in many observation data. Overfitting of rare input situations is avoided this way.
The confident normalized hit rate can be used to generate a small rule base covering the relevant part of the observation data. Scarce situations caused by random effects are not represented by the rule base.

The t–test also is a test on hypotheses about the observed data, but the mean of the output data given the premise conditions of a rule is tested instead of the hit rate or probability of its conclusion part. If the constrained output mean is significantly different from the mean of all output data, the premise is considered to be relevant and a rule is generated. An accepted rule is rated by the relative standard deviation of the constrained output value y_p.
In current experiments, the mean $\bar{y}$ was removed from the observed output data when applying the t–test. A solution for modeling a nonzero mean value $\bar{y}$ of the output data y is under investigation.

The t–test is expected to be the best solution for modeling and predicting noisy data when the mean output error is to be minimized. It is restricted, however, to metric output variables.

The new testing strategies are currently studied in test problems. The studies concentrate on the influence of the confidence level α and the sample size. In the study several examples of one dimensional test data are modeled with the Fuzzy–ROSA method using the different test strategies at different confidence levels. The results will be presented in detail in [10].

Further research is also done on the use of the new strategies in the field of load prediction. In this application the total demand for electric power in a control area is predicted using the Fuzzy-Rosa method.

References

[1] H. Kiendl and M. Krabs. Ein Verfahren zur Generierung regelbasierter Modelle für dynamische Systeme. *Automatisierungstechnik*, 37/Heft 11 : pages 423–430, 1989.

[2] A. Krone and H. Kiendl. Automatic Generation of Positive and Negative Rules for Two-Way Fuzzy Controllers. In *Second EUFIT (European Conference on Intelligent Techniques and Soft Computing* , pages 438–447, Aachen, 1994.

[3] H. Kiendl, M. Krabs, and M. Fritsch. Rule-Based Modelling of Dynamical Systems. In *Analysis and Control of Industrial Processes*, pages 217–231, Vieweg-Verlag, Braunschweig, 1991.

[4] A. Krone and H. Kiendl. Rule–Based Decision Analysis with Fuzzy–Rosa Method. In Felix, R. (ed.), *EFDAN(European Workshop on Fuzzy Decision Analysis for Management, Planning and Optimization)* , pages 109–114, 1996.

[5] A. Krone. Advanced rule reduction concepts for optimizing efficiency of knowledge extraction. In *Fourth EUFIT (European Congress on Intelligent Techniques and Soft Computing)*, pages 919–923, Aachen, 1996.

[6] A. Krone, Ch. Frenck, and O. Russak. Design of a Fuzzy Controller for an Alkoxylation Process using the ROSA–Method for Automatic Rule Generation. In *Third EUFIT (European Congress on Intelligent Techniques and Soft Computing)*, pages 760–764, Aachen, 1995.

[7] M. Krabs. Das ROSA–Verfahren zur Modellierung dynamischer Systeme durch Regeln mit statistischer Relevanzbewertung, PhD thesis, 1994.

[8] H. Kiendl. *Fuzzy Control methodenorientiert.* Oldenbourg, München, Wien, 1997.

[9] P.R. Cohen. *Empirical Methods for Artificial Intelligence.* MIT Press, 1995.

[10] T. Slawinski and H. Jessen. Vergleichende Untersuchung und Anwendungen der verschiedenen Regeltest- und Bewertungsstrategien im Fuzzy–ROSA–Verfahren. In *Reihe Computational Intelligence, Collaborative Research Center 531 (Design and Management of Complex Technical Processes and Systems by means of Computational Intelligence Methods)* (to appear).