Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

ELSEVIER International Journal of Approximate Reasoning 35 (2004) 195–202
www.elsevier.com/locate/ijar

Influence of fuzzy norms and other heuristics on "Mixed fuzzy rule formation"

Thomas R. Gabriel, Michael R. Berthold *

Tripos Inc., Data Analysis Research Lab., 601 Gateway Blvd., Suite 720, South San Francisco, CA 94080, USA

Department of Computer and Information Science, University of Konstanz, P.O. Box M712, 78457 Konstanz, Germany

Received 1 June 2003; accepted 1 October 2003

Abstract

In *Mixed Fuzzy Rule Formation* [Int. J. Approx. Reason. 32 (2003) 67] a method to extract mixed fuzzy rules from data was introduced. The underlying algorithm's performance is influenced by the choice of fuzzy *t*-norm and *t*-conorm, and a heuristic to avoid conflicts between patterns and rules of different classes throughout training. In the following addendum to [Int. J. Approx. Reason. 32 (2003) 67], we discuss in more depth how these parameters affect the generalization performance of the resulting fuzzy rule models.

1. Mixed fuzzy rule formation

The training method described in [1] is based on an iterative algorithm. During each learning epoch, i.e. presentation of all training patterns, new fuzzy rules are introduced when necessary and existing ones are adjusted whenever a conflict occurs. For each pattern three main steps are executed. Firstly, if a new training pattern lies inside the support-region of an existing fuzzy rule of the

* Corresponding author. Address: Department of Computer and Information Science, University of Konstanz, P.O. Box M712, 78457 Konstanz, Germany.

E-mail addresses: gabriel@inf.uni-konstanz.de (T.R. Gabriel), berthold@inf.uni-konstanz.de (M.R. Berthold).

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2003.10.004

correct class, its core-region is extended in order to cover the new pattern (*cover*). Secondly, if the new pattern is not yet covered, a new fuzzy rule of the correct class is introduced (*commit*). The new example is assigned to its core, whereas the support-region is initialized "infinite", that is, the new fuzzy rule covers the entire domain. Lastly, if a new pattern is incorrectly covered by an existing fuzzy rule, the fuzzy point's support-region is reduced so that the conflict is avoided (*shrink*). This heuristic for conflict avoidance aims to minimize the loss in volume [1]. In Section 2 three different heuristics to determine the loss in volume are compared in more detail. As discussed in [1], the algorithm terminates after only few iterations over the set of example patterns. The resulting set of fuzzy rules can then be used to classify new patterns by computing the overall fuzzy membership degree. The accumulated membership degrees over all input dimensions and across multiple rules are calculated using fuzzy t-norm resp. t-conorm. Again, [1] does not discuss different fuzzy norms, thus we present some choices in Section 3 in more detail and show how they can affect the classification accuracy.

2. Shrink heuristics

As mentioned above, the training procedure relies on a heuristic which affects the strategy to avoid conflicts. We have several different choices for this conflict avoidance heuristic. One common approach is to shrink the fuzzy rule in dimension i $(1 \leqslant i \leqslant n)$ that minimizes the loss in volume:

$$i_{\min} = \arg\min_{i=1,\dots,n}\{V_i\}.$$

The loss in volume V_i of a fuzzy rule $\mathbf{R}$ (using trapezoid membership functions with parameters $\langle a, b, c, d\rangle$ where (a, b) and (c, d) bound the support-region, and $[b, c]$ the fuzzy rule's core-region) is then:

$$V_i = d_i^*(\vec{x}, \mathbf{R}) \cdot \prod_{j=1, j\neq i}^{n} d_j^\times(\vec{x}, \mathbf{R}),$$

where $d_i^*(\cdot)$ $(1 \leqslant i \leqslant n)$ is the distance between example pattern $\vec{x}$ and the border (core- or support-region) of a fuzzy rule $\mathbf{R}$ in dimension i, and $d_j^\times(\cdot)$ $(1 \leqslant j \leqslant n)$ indicates the distance to fuzzy rule $\mathbf{R}$ in dimension j. Later in this section, these two functions are defined more precisely. Furthermore, the loss in volume is normalized with respect to the overall volume:

$$V_i^{\text{norm}} = d_i^*(\vec{x}, \mathbf{R}) \cdot \frac{\prod_{j=1, j\neq i}^{n} d_j^\times(\vec{x}, \mathbf{R})}{\prod_{j=1}^{n} d_j^\times(\vec{x}, \mathbf{R})} = d_i^*(\vec{x}, \mathbf{R}) \cdot d_i^\times(\vec{x}, \mathbf{R}).$$

That is, the computation of this loss in volume can be simplified because only the shrunken dimension has to be considered. However, the losses in volume in the core- and support-region still have to be treated separately.

If the conflict occurs in the support-region ($a < x < b \vee c < x < d$), the volume loss function $d_i^*(\cdot)$ between the outer left resp. right (marked by the initial vector v_i) border of the support- and core-region is:

$$d_i^*(\vec{x}, \mathbf{R}) = \begin{cases} x_i - a_i, & x_i \leqslant v_i, \\ d_i - x_i, & \text{otherwise.} \end{cases}$$

The function $d_i^\times(\cdot)$ weighting the loss in volume still needs to be defined. We introduce three alternative shrink heuristics:

Rule-based shrink: weights the loss in volume with respect to the entire fuzzy rule spread:

$$d_i^\times(\vec{x}, \mathbf{R}) = d_i - a_i.$$

Anchor-based shrink: weights the loss in volume with respect to the distance between the initial vector (anchor) and the border of the fuzzy rule's support-region:

$$d_i^\times(\vec{x}, \mathbf{R}) = \begin{cases} v_i - a_i, & x_i \leqslant v_i, \\ d_i - v_i, & \text{otherwise.} \end{cases}$$

Area-based shrink: weights the loss in volume with respect to the distance between the border of the fuzzy rule's support- and core-region:

$$d_i^\times(\vec{x}, \mathbf{R}) = \begin{cases} b_i - a_i, & x_i \leqslant v_i, \\ d_i - c_i, & \text{otherwise.} \end{cases}$$

If the conflict is part of the core-region ($b \leqslant x \leqslant c$), the function $d_i^*(\cdot)$ is:

$$d_i^*(\vec{x}, \mathbf{R}) = \begin{cases} x_i - b_i, & x_i \leqslant v_i, \\ c_i - x_i, & \text{otherwise,} \end{cases}$$

and we have two choices in order to compute the function $d_i^\times(\cdot)$.

Rule-/area-based shrink: weights the loss in volume with respect to the spread of the fuzzy rule's core-region:

$$d_i^\times(\vec{x}, \mathbf{R}) = c_i - b_i.$$

Anchor-based shrink: weights the loss in volume with respect to the distance between the border of the core-region and the fuzzy rule's anchor (initial vector):

$$d_i^{\times}(\vec{x}, \mathbf{R}) = \begin{cases} v_i - b_i, & x_i \leqslant v_i, \\ c_i - v_i, & \text{otherwise.} \end{cases}$$

For our tests these three shrink heuristics are used for evaluation. The next section discusses the results on benchmark data sets in more detail.

3. Fuzzy norms

The algorithm described in [1] constructs a set of fuzzy rules that can be used to classify new example instances of unknown class. The one-dimensional rule antecedents are combined using a fuzzy t-norm and the degrees of fulfillment of all rules of one class are combined using a t-conorm, resulting in a final degree of membership for each class. The choice of these norms have a noticeable influence on the classification outcome.

The most popular choice for these fuzzy norms was introduced by Zadeh in [5]:

$$\top(\mu(x), \mu(y)) = \min\{\mu(x), \mu(y)\},$$
$$\bot(\mu(x), \mu(y)) = \max\{\mu(x), \mu(y)\},$$

where $\mu(\cdot)$ is the degree of membership of a fuzzy rule, $\top(\cdot)$ (t-norm) the fuzzy operator for the conjunction, and $\bot(\cdot)$ (t-conorm) the operator for the disjunction. This so-called minimum/maximum norm represents the most optimistic resp. most pessimistic choice for these operators. Other common choices are the product norm:

$$\top(\mu(x), \mu(y)) = \mu(x) \cdot \mu(y),$$
$$\bot(\mu(x), \mu(y)) = \mu(x) + \mu(y) - \mu(x) \cdot \mu(y),$$

and the Łukasiewicz norm [2]:

$$\top(\mu(x), \mu(y)) = \max\{0, \mu(x) + \mu(y) - 1\},$$
$$\bot(\mu(x), \mu(y)) = \min\{1, \mu(x) + \mu(y)\}.$$

In [4] an entire family of fuzzy norms called Yager norm is defined as:

$$\top_\omega(\mu(x), \mu(y)) = 1 - \min\left\{1, [(1 - \mu(x))^\omega + (1 - \mu(y))^\omega]^{\frac{1}{\omega}}\right\},$$
$$\bot_\omega(\mu(x), \mu(y)) = \min\left\{1, [\mu(x)^\omega + \mu(y)^\omega]^{\frac{1}{\omega}}\right\}, \quad \omega \in]0, \infty[.$$

The definitions above are probably the most well-known choices for fuzzy t-norms and t-conorms. We focus our experiments using the Yager norm with $\omega = 2^{-1}$ and $\omega = 2$, in addition to the minimum/maximum, product, and

Table 1
Used data sets along with the number of features, classes, train, and test data

Data set	# features	# classes	# train data	# test data
Diabetes	8	2	768	12-fold
Aust. Cred.	14	2	690	10-fold
Vehicle	18	4	846	9-fold
Segment	11	7	2.310	10-fold
Shuttle	9	7	43.500	14.500
SatImage	36	6	4.435	2.000
DNA	240	3	2.000	1.186
Letter	16	26	15.000	5.000

Łukasiewicz norm in order to evaluate their influence on the classification performance using the same benchmark data sets as in [1].

4. Experimental results

The evaluation of the proposed methodology is conducted using eight data sets from the StatLog-Project [3]. [1] Table 1 shows the properties of the used data sets. All sets are divided into train and test data (see last two columns). On the first four data sets we perform *n*-fold cross validation following [3] due to the small number of examples (the last column shows the number of folds).

As mentioned before, the classification accuracy is compared using five different fuzzy norms—minimum/maximum, product, $\text{Yager}_{1/2}$, Łukasiewicz, and Yager_2 and also three shrink heuristics for conflict avoidance—rule-, anchor-, and area-based shrink. Tables 2–4 summarize the error rates in percent for each data set. The tables are grouped by shrink heuristic first and fuzzy norm second to compare the parameters' influences individually.

Fig. 1 shows a graphical summary grouped by shrink heuristics. It is obvious to see that most strategies only have a weak influence on the model's generalization performance for the different choices of fuzzy norms. The $\text{Yager}_{1/2}$ norm is the outlier in this case, always providing results which are substantially worse than the others. Better results can be achieved using the minimum/maximum as well as product norm. In addition, these two norms seem to be more stable on the data sets used here. The Yager_2 norm reaches similar results, the best in the anchor- and area-based group. The Łukasiewicz norm achieves good results in comparison to the other fuzzy norms but always slightly worse than the error average.

[1] The remaining 12 data sets are either not suitable for the underlying FRL algorithm (contain categorical variables) or were not available for download.

Table 2
Rule-based shrink heuristic along with fuzzy norms

Data set	Min/max	Product	$\text{Yager}_{1/2}$	Łuka	Yager_2
Diabetes	29.03	29.43	30.99	29.43	26.30
Aust. Cred.	17.10	17.10	17.10	16.81	16.81
Vehicle	38.18	37.71	46.10	39.72	36.88
Segment	7.79	7.92	13.81	9.61	8.01
Shuttle	0.08	0.08	0.06	0.07	0.09
SatImage	16.50	16.55	26.30	20.20	17.75
DNA	32.63	32.63	36.59	36.93	32.29
Letter	24.28	24.32	29.47	25.79	23.24

Table 3
Anchor-based shrink heuristic along with fuzzy norms

Data set	Min/max	Product	$\text{Yager}_{1/2}$	Łuka	Yager_2
Diabetes	28.78	29.17	30.21	27.21	26.82
Aust. Cred.	17.82	17.68	17.97	17.25	17.25
Vehicle	32.51	32.15	45.51	36.67	29.91
Segment	4.55	4.59	10.61	4.85	4.16
Shuttle	0.06	0.06	0.08	0.07	0.07
SatImage	14.20	13.90	29.65	20.95	14.10
DNA	32.63	32.72	36.68	36.51	32.88
Letter	14.44	14.20	20.18	16.46	14.12

Table 4
Area-based shrink heuristic along with fuzzy norms

Data set	Min/max	Product	$\text{Yager}_{1/2}$	Łuka	Yager_2
Diabetes	32.29	31.90	28.91	29.56	27.21
Aust. Cred.	18.84	18.55	16.67	18.83	17.10
Vehicle	32.98	32.98	46.34	38.06	31.56
Segment	3.90	3.85	9.57	4.85	4.16
Shuttle	0.06	0.06	0.08	0.06	0.06
SatImage	13.75	13.80	24.55	19.50	14.75
DNA	32.72	32.97	36.59	36.93	31.96
Letter	14.36	14.28	20.18	15.66	14.92

Fig. 2 summarizes the results for the three shrink heuristics grouped by fuzzy norms. The graphic shows that the choice of fuzzy norm only has a small influence on the performance of the generated model. But it is interesting to see that the rule-based shrink heuristic consistently delivers worse results than any of the other strategies. The anchor- and area-based heuristics provide almost same results except for the $\text{Yager}_{1/2}$ norm (as already discussed before).

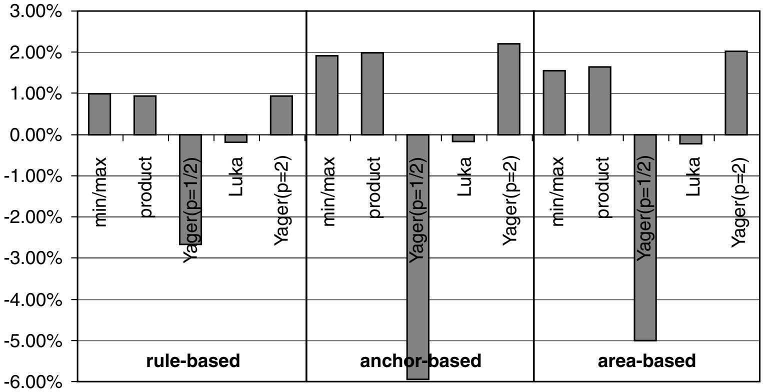

Fig. 1. Deviation from the average error rate grouped by the three shrink heuristics—rule-, anchor-, and area-based.

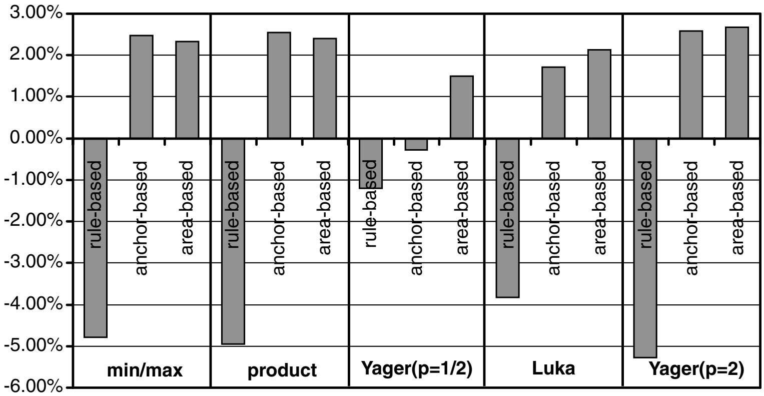

Fig. 2. Deviation from the average error rate grouped by the five fuzzy norms—minimum/maximum, product, $\text{Yager}_{1/2}$, Łukasiewicz, and Yager_2.

5. Conclusion

In this addendum to [1], we showed how the choice of different fuzzy norms affects the generalization performance of the resulting rule systems considerably. We also demonstrated how various heuristics to adjust rules for conflict avoidance with new training instances affect the performance of the rule system. For most choices, the algorithm behaves well as long as Yager fuzzy norms and rule-based shrink heuristics are avoided. In general this addendum illustrates that the choice of conflict avoidance heuristic and fuzzy norm can affect the final classification performance of the resulting fuzzy rule model substantially. However, if the most drastic choices are avoided, the outcome is reasonably independent of the task at hand.

References

[1] M.R. Berthold, Mixed fuzzy rule formation, Int. J. Approx. Reason. (IJAR) 32 (2003) 67–84.
[2] J. Łukasiewicz, Selected Works—Studies in Logic and the Foundations of Mathematics, North-Holland, Amsterdam, 1970.
[3] D. Michie, D.J. Spiegelhalter, C.C. Taylor (Eds.), Machine Learning, Neural and Statistical Classification, Ellis Horwood Limited, 1994.
[4] R.R. Yager, S. Ovchinnikov, R.M. Tong, H.T. Ngugen (Eds.), Fuzzy Sets and Applications, Wiley, New York, 1987.
[5] L.A. Zadeh, Fuzzy sets, Inform. Control 8 (1965) 338–353.