Automatic design of interpretable membership functions

Ralf Mikut, Jens Jäkel, Lutz Gröll

Forschungszentrum Karlsruhe GmbH
Institute for Applied Computer Science (IAI)
P.O. Box 3640, D-76021 Karlsruhe, Germany
Phone: +49-7247-825731, Fax: +49-7247-825785
E-Mail: {mikut,jaekel,groell}@iai.fzk.de

1 Motivation

Most approaches for the data-based design of fuzzy rule-based classifiers (e. g. [1–4]) require a manual definition of membership functions (MBFs). In large applications, this task is very difficult and time-consuming for human experts. As a practical application, a possible scenario for a manual design of membership functions in a medical expert system to support gait analysis is demonstrated (Section 2).

Therefore, in the field of fuzzy modelling heuristic approaches like equidistant MBFs, clustering [5] or parameter optimization [6, 7] are applied for an automatic definition of MBFs. The drawbacks of these approaches (e. g. lack of interpretability, explanation of output classes) become substantial in the case of high input dimensionality.

The principal idea of this paper is to construct an appropriate criterion for evaluating partitions or MBFs and to search for the optimal value. The algorithm bases on information-theoretical measures [8] and additional modifications to guarantee the interpretability and practical acceptance of the solutions. Because information-theoretical measures assume a discrete partition of input and output variables, new approaches to represent fuzzy membership values as a possibility distribution are introduced.

The aims of this paper are

- to give a short overview about information-theoretical measures (Section 3),
- to propose a new approach to compute them for a possibility distribution assuming a standard fuzzy partition (Section 4),
- to present an automatic design method using a combined criterion for the evaluation of fuzzy partitions and to show the results for the practical application from Section 2 (Section 5).

2 Classification of gait phases

Usually, the interpretation and classification of medical data depends on the experience of highly specialized clinicians. As an example, the instrumented gait analysis [9] extracts kinematic and kinetic data of human movements from video records and force measurements. Its aim is to improve clinical decision making and therapy. Typical diseases are paraplegia (see e. g. [10]) and cerebral palsy (see e. g. [11]). In addition, gait analysis is used for the design of prosthesis (see e. g. [12]) and for the development of walking robots. Up to now, these decisions are made on subjective criteria. The aim of a joint project of the Department of Orthopaedic Surgery, University of Heidelberg, and the Forschungszentrum Karlsruhe is to support clinicians by computer aided gait analysis. The used methods are fuzzy logic and multivariate statistics for evaluation, quantification and extension of expert knowledge.

As a first application, the categorisation of kinematic and kinetic data into seven specified phases of the gait cycle (output classes) is demonstrated (see Fig. 1). As input variables, 174 features are available. The design of fuzzy classifiers to explain these output classes by fuzzy rules is demonstrated in [13,14]. The dimensionality of the input space requires a fully automatic MBF design. Fig. 2 shows the difficulty of defining membership functions for a relevant feature, the flexion angle of the right knee in the sagittal plane. To separate the 5th and the 6th output class, it is necessary to get a partition border of appr. 45°. To divide the first four classes from the 5th class, a further border should be placed between 15° and 25°. All other decisions are very difficult. In this case it is important to divide the remaining regions because a combination with further features could solve

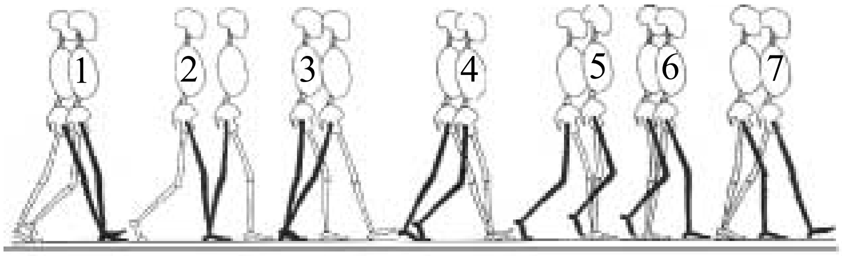
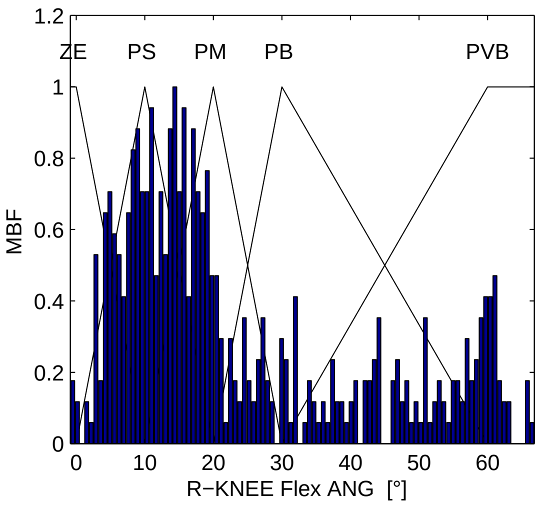


Figure 1: Gait phases (1: Initial Contact und Loading Response, 2: Mid Stance, 3: Terminal Stance, 4: Pre Swing, 5: Initial Swing, 6: Mid Swing, 7: Terminal Swing [9])



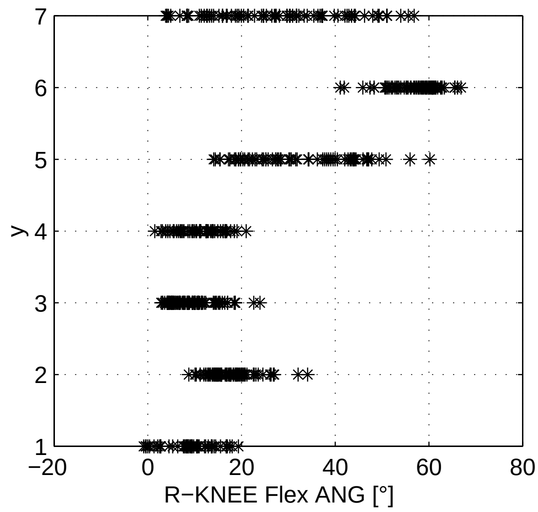


Figure 2: Membership function for the right knee flexion (P positive, S small, M medium, B big, ZE zero, V very) with its histogram (left) and distribution of output classes (y-axis, right) depending on the right knee flexion (x-axis,right)

the conflicts. In the region with a higher density of samples (between 5° and 20°), the density of membership functions should also be higher.

All parameters of an expert solution (Fig. 2, left, parameters 0°, 10°, 20°, 30° and 60°) have rounded values and meet the semantics of the term names (e. g. zero). In the next sections, this manual design approach will be formalized to enable its automation.

3 Information-theoretical measures

In the considered problem, each input variable x_l has m_l linguistic terms $A_{l,i}$ and the output variable y has m_y linguistic terms B_j.

The well-known information-theoretical measures (see e. g. [8]) can be applied as a tool to evaluate the fitness of discrete partitions of input variables (measurement data) and output variables (classification results). They base actually on class probabilities and provide an abstract measure for relationships between input ($A_{l,i}$) and output classes (B_j).

Fig. 3 shows different measures for the relationship between a feature x_l (measurement value) and the output value y. The input entropy $H(x_l)$, the output entropy $H(y)$ and the joint entropy $H(x_l, y)$ can be computed by

$$H(x_l) = -\sum_{i=1}^{m_l} p(x_l \in A_{l,i}) \cdot \log_2 p(x_l \in A_{l,i}) \tag{1}$$

$$H(y) = -\sum_{j=1}^{m_y} p(y \in B_j) \cdot \log_2 p(y \in B_j) \tag{2}$$

$$H(x_l, y) = -\sum_{i=1}^{m_l} \sum_{j=1}^{m_y} p(x_l \in A_{l,i} \wedge y \in B_j) \cdot \log_2 p(x_l \in A_{l,i} \wedge y \in B_j) \tag{3}$$

using the observed class frequencies as estimation for the probabilities:

$$\hat{p}(x_l \in A_{l,i}) = n(x_l \in A_{l,i})/N \tag{4}$$

$$\hat{p}(y \in B_j) = n(y \in B_j)/N \tag{5}$$

$$\hat{p}(x_l \in A_{l,i} \wedge y \in B_j) = n(x_l \in A_{l,i} \wedge y \in B_j)/N. \tag{6}$$

Here, N is the total number of samples, n the number of samples, where x_l has the value $A_{l,i}$ ($n(x_l \in A_{l,i})$), y has the value B_j ($n(y \in B_j)$), and where the conjunction of both conditions is true ($n(x_l \in A_{l,i} \wedge y \in B_j)$).

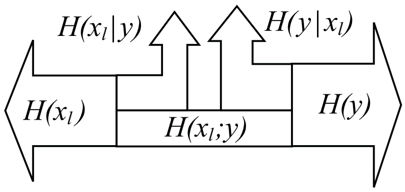


Fig. 3: Entropy measures

The mutual information $H(x_l; y)$ characterizes the information flow between input and output variable. Often but not always, it is a sign for a causal dependency between these variables. The irrelevance $H(y|x_l)$ (conditional entropy of y given x_l) and the equivocation $H(x_l|y)$ (conditional entropy of x_l given y) measure the parts of the output and input variable which can not be explained by the other variable. The equations

$$H(x_l; y) = H(x_l) + H(y) - H(x_l, y) \tag{7}$$
$$H(y|x_l) = H(x_l, y) - H(x_l) \tag{8}$$
$$H(x_l|y) = H(x_l, y) - H(y). \tag{9}$$

describe their relations.

Because the estimation of probabilities leads in all information theoretical measures to a large bias for a small number of samples N (input and output entropy will be estimated to small, the mutual information to high), additional corrections are used [15].

4 Fuzzy information-theoretical measures

For fuzzy systems, the same equations (1)–(3) and (7)–(9) can be used to analyze possibility distributions. All equations actually base on mutual exclusive classes with membership values of 0 or 1. Now, a solution to transform fuzzy membership values of linguistic terms *between* 0 and 1 into the estimated class frequencies has to be found instead of (4)–(6).

In a matrix description, the membership values of an input variable x_l can be written as $\boldsymbol{\mu}(x_l)$ with the elements $\mu_{A_{l,i}}(x_l[k])$ in $i = 1, \cdots, m_l$ rows (linguistic terms) and $k = 1, \cdots, N$ columns (samples). Analogously, $\boldsymbol{\mu}(y)$ contains the elements $\mu_{B_j}(y[k])$ in $j = 1, \cdots, m_y$ rows (linguistic terms) and N columns (samples).

There are three different approaches to represent the fuzziness:

1. discretization by using the linguistic term with the maximal membership value,
2. summing up all membership values [16]:

$$\hat{p}(x_l \in A_{l,i}) = \frac{1}{N} \sum_{k=1}^{N} \mu_{A_{l,i}}(x_l[k]) \tag{10}$$
$$\hat{p}(y \in B_j) = \frac{1}{N} \sum_{k=1}^{N} \mu_{B_j}(y[k]) \tag{11}$$
$$\hat{p}(x_l \in A_{l,i} \wedge y \in B_j) = \frac{1}{N} \sum_{k=1}^{N} \mu_{A_{l,i}}(x_l[k]) \cdot \mu_{B_j}(y[k]), \tag{12}$$

3. to use only (10)–(11) and to compute $\hat{p}(x_l \in A_{l,i} \wedge y \in B_j)$ by

$$\hat{p}(x_l \in A_{l,i} \wedge y \in B_j) = \hat{p}_{j,i}(y|x_l) \cdot \hat{p}(x_l \in A_{l,i}). \tag{13}$$

   $\hat{p}_{j,i}(y|x_l)$ is the element (j, i) in a matrix $\hat{\boldsymbol{p}}(y|x_l)$ containing the conditional class frequencies with given classes of x_l. As a consequence, this element (j, i) is the estimation for the relative class frequency of $y = B_j$ under the condition $x_l = A_{l,i}$. To get this matrix, a constrained optimization problem

$$Q(\hat{\boldsymbol{p}}(y|x_l)) = \tfrac{1}{2} \|\hat{\boldsymbol{p}}(y|x_l)\boldsymbol{\mu}(x_l) - \boldsymbol{\mu}(y)\|_F^2 \rightarrow \underset{\hat{\boldsymbol{p}}(y|x_l)}{\mathrm{Min}}, \tag{14}$$

has to be solved subject to

$$\hat{\boldsymbol{p}}(y|x_l) \geq \mathbf{0}_{m_y \times m_l} \tag{15}$$

$$\mathbf{1}_{m_y}^T \hat{\boldsymbol{p}}(y|x_l) = \mathbf{1}_{m_l}^T. \tag{16}$$

These constraints assure that $\hat{\boldsymbol{p}}(y|x_l)$ only contains non-negative elements (15) and its column sums are one (16). Corresponding to a probability interpretation, these contraints guarantee that there do not exist negative probabilities and that the sum of all probabilities is one. The solution is

$$\hat{\boldsymbol{p}}(y|x_l)_{opt} = \underbrace{\left(\boldsymbol{I}_{m_y} - \tfrac{1}{m_y}\mathbf{1}_{m_y \times m_y}\right)}_{\boldsymbol{Z}_{m_y}} \underbrace{\boldsymbol{\mu}(y)\boldsymbol{\mu}(x_l)^T \left(\boldsymbol{\mu}(x_l)\boldsymbol{\mu}(x_l)^T\right)^{-1}}_{\hat{\boldsymbol{p}}(y|x_l)_{LS}}$$

$$+ \tfrac{1}{m_y}\mathbf{1}_{m_y \times m_l} + \underbrace{\left(\boldsymbol{I}_{m_y} - \tfrac{1}{m_y}\mathbf{1}_{m_y \times m_y}\right)}_{\boldsymbol{Z}_{m_y}} \boldsymbol{L}^* \left(\boldsymbol{\mu}(x_l)\boldsymbol{\mu}(x_l)^T\right)^{-1} \tag{17}$$

with the Lagrangian matrix $\boldsymbol{L}^*$, the unconstrained least square solution $\hat{\boldsymbol{p}}(y|x_l)_{LS}$, a vector of ones $\mathbf{1}$ and the identity matrix $\boldsymbol{I}$. As well known (e. g. [17]) for a strictly convex programming problem with linear constraints a solution $\hat{\boldsymbol{p}}(y|x_l)_{opt}$ which meets the Kuhn-Tucker conditions is unique.

In an illustrative example, a data set with $N = 76$ samples

$$\boldsymbol{x}_l = \begin{pmatrix} 0.00 & 0.01 & \cdots & 0.50 & 0.52 & \cdots & 0.98 & 1.00 \end{pmatrix} \tag{18}$$

$$\boldsymbol{y} = \begin{pmatrix} 1.00 & 0.99 & \cdots & 0.50 & 0.48 & \cdots & 0.01 & 0.00 \end{pmatrix} \tag{19}$$

is given. It is easy to see that two trapezoidal input MBFs with the parameters $a_{l,1}^R = 0$ and $a_{l,2}^L = 1$ (see Section 5 and Fig. 4) are optimal in combination with identical output MBFs leading to the desired result

$$\boldsymbol{\mu}(x_l) = \begin{pmatrix} 0.00 & 0.01 & \cdots & 0.50 & 0.52 & \cdots & 0.98 & 1.00 \\ 1.00 & 0.99 & \cdots & 0.50 & 0.48 & \cdots & 0.02 & 0.00 \end{pmatrix} \tag{20}$$

$$\boldsymbol{\mu}(y) = \begin{pmatrix} 1.00 & 0.99 & \cdots & 0.50 & 0.48 & \cdots & 0.02 & 0.00 \\ 0.00 & 0.01 & \cdots & 0.50 & 0.52 & \cdots & 0.98 & 1.00 \end{pmatrix}. \tag{21}$$

For this choice, the first and the third approach to represent fuzziness from Section 3 both leads to

$$\frac{H(x_l;y)}{H(y)} = 0.96, \quad \hat{\boldsymbol{p}}(y|x_l) = \begin{pmatrix} 0.00 & 1.00 \\ 1.00 & 0.00 \end{pmatrix}, \quad \hat{\boldsymbol{p}}(y \wedge x_l) = \begin{pmatrix} 0.00 & 0.42 \\ 0.58 & 0.00 \end{pmatrix}, \tag{22}$$

but the second approach dramatically underestimates the weighted mutual information:

$$\frac{H(x_l;y)}{H(y)} = 0.06, \quad \hat{\boldsymbol{p}}(y|x_l) = \begin{pmatrix} 0.28 & 0.60 \\ 0.72 & 0.40 \end{pmatrix}, \quad \hat{\boldsymbol{p}}(y \wedge x_l) = \begin{pmatrix} 0.16 & 0.26 \\ 0.42 & 0.16 \end{pmatrix}. \tag{23}$$

The consequences of this underestimation will be shown in the next section.

5 Design of membership functions

This part describes the design of the membership functions for the input variables of the fuzzy system. The resulting optimization problem will be decomposed in sub-problems with m_l parameters for each input variable. Actually, the approach can also be transfered to output membership functions but the computation effort here increases remarkably. In addition, the output membership functions should not be changed between the sub-problems of the different input variables.

Using a parametric approach, the i-th trapezoidal input membership function of the l-th feature is completely characterized by the positions of the left and right maximum ($a^L_{l,i}$ and $a^R_{l,i}$) if only adjoining MBFs overlap and the sum of all membership values is one. Triangular MBFs hold $a_{l,i} = a^L_{l,i} = a^R_{l,i}$. As a consequence, a fuzzy partition with trapezoidal MBFs for the left and right term and triangular MBFs in the middle is defined by a parameter vector $\boldsymbol{a}$ with m_l parameters. For the illustrative example from Section 4 in the figure (top), the parameter vector is $\boldsymbol{a} = \begin{pmatrix} a^R_{l,1} & a^L_{l,2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \end{pmatrix}$.

From each fuzzy partition (Fig. 4, top), two different crisp partitions directly follow: the partition by the maximal membership degree (Fig. 4 in the middle, m_l terms) and the partition between the parameters of adjoining terms, left of the smallest and right of largest parameter (Fig. 4 bottom, $m_l + 1$ intervals $I_{l,i}$). For the placement of MBFs, different contradictory aims exist:

- separation of output classes (including to avoid the rejection of unfrequent output classes),
- similar frequency of examples in all terms,
- avoidance of very small distances between the parameters of adjoining MBFs and
- interpretability according to the semantics of terms (e. g. $a^R_{l,1} = 0.00$ instead of $a^R_{l,1} = 0.07$ for the term zero).

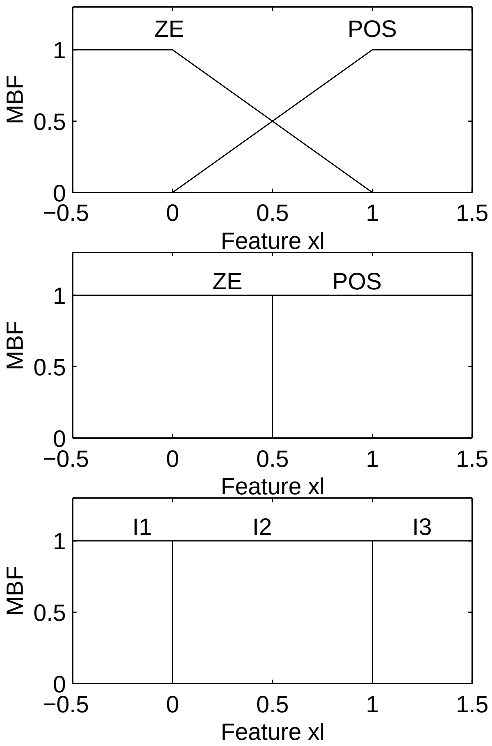


Fig. 4: Fuzzy and crisp partitions

For each of these aims, a criterion can be constructed: A good separation of output classes is reached by maximizing mutual information (7). The maximization of the input entropy (1) leads to equal frequencies of examples in all terms. Similarly, a maximization of the interval input entropy

$$H_I(x_l) = -\sum_{i=1}^{m_l+1} p(x_l \in I_{l,i}) \cdot \log_2 p(x_l \in I_{l,i}) \qquad (24)$$

avoids very small distances between the parameters of adjoining MBFs. Interpretable parameter values can be obtained if preferable values (e. g. 0.00) get a higher value of the interpretability criterion compared to other values.

The interpretability criterion I divides possible parameters into different interpretability classes (very high, high, medium, none). It has large positive values at preferred parameters, the value zero for unacceptable values and smaller positive values for compromise parameters (Table 1). With this criterion, the preferred values near to the optimal values will be chosen. A special algorithm adapts the interpretability classes to the minimal and maximal values of the input variable. The interpretability criterion I, the information-

Class	Parameter $a_{l,i}$	Interpretability criterion I
I	0	very high (0.15)
II	-1; 1; 2	high (0.10)
III	-0.5; -0.4 ; $\cdots$; 1.5	medium (0.05)
IV	e. g. -0.49; -0.4855; $\cdots$	none (0.00)

Table 1: Interpretability criterion I and its values for different parameters for MBFs with $x_l \in [-0.5; 1.5]$

theoretical measures $H(x_l; y)$ and $H_I(x_l)$ depend on the parameter vector of the input MBFs $\boldsymbol{a}_l$. The optimal parameter vector maximizes a compromise criterion

$$Q = \alpha \frac{H(x_l; y)}{H(y)} + (1 - \alpha) \frac{H_I(x_l)}{H_{I,max}(x_l)} + \beta I \rightarrow \max_{\boldsymbol{a}_l} . \tag{25}$$

The parameter α controls the preference between an optimal separation of output classes ($\alpha = 1$) and similar class frequencies ($\alpha = 0$). This compromise will be influenced by parameter β switching softly between optimal parameters for the compromise ($\beta = 0$) and interpretable parameters ($\beta = 1$). The scaling factor $H_{I,max}(x_l)$ is the maximal input entropy for the given number of intervals.

The computational effort is high because each parameter test requires a new fuzzification. To minimize this effort, a suboptimal algorithm will be used. The algorithms starts with the minimal value of x_l in the data set as parameter $a^R_{l,1}$, its maximal value as parameter a^L_{l,m_l} and with parameters guaranteeing an equal number of samples for all terms. In this initial value of the parameter vector $\boldsymbol{a}_l$, the parameters will be deleted step by step and replaced by one new parameter maximizing (25).

The results of (25) depends crucially on the fuzzy representation from Section 4. For the illustrative example in the same section, the underestimation of mutual information with the sum approach in (23) causes a misplacement of MBF parameters (e. g. $a^R_{l,1} = 0.4$ in the Fig. 5 right). In other problems, the first approach (maximal membership value) overestimates the fitness of membership function and it fails in the fine-tuning of MBFs. The constrained optimization approach finds the optimal parameters $a^R_{l,1} = 0.0$ and $a^L_{l,2} = 1.0$ (Fig. 5 left). In general, it seems to be the best in these three approaches.

The decisions of different sub-criteria for the practical application from Section 2 are displayed in Table 2. The criteria C, D and E find the term borders 47.46, 46.60, 45.00 and 22.46, 18.15 and 25.00 (mean values between adjoining parameters) but use the remaining parameters in different ways. The criterion C only maximizes the mutual information and leads to a very small distance between the parameters $a_{1,3}$ and $a_{1,4}$. A good compromise is the decision with criterion E (Q with I) which gives exact the same result as the expert solution is Fig. 2 (left).

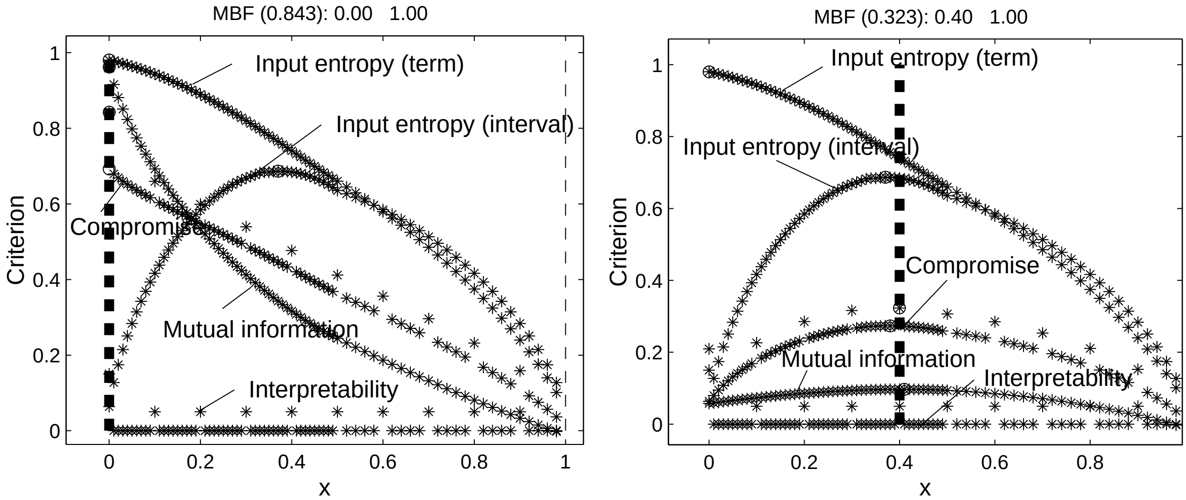


Figure 5: Optimization of the parameter $a_{1,1}$ for different fuzzy representations - left: constrained optimization (13)–(16), right: summation (12)

The replacement of the parameter $a^L_{1,5}$ using the different sub-criteria in (25) shows Fig. 6 - the last step of the optimization with the criterion E. Both input entropies and the output entropy are displayed in their normalized forms.

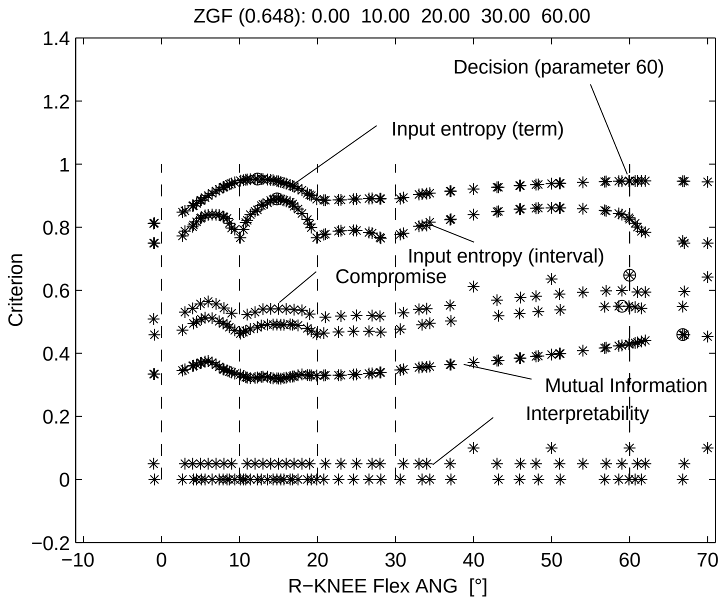


Figure 6: Values of different sub-criteria in the optimization of the 5th membership function (PVB, $a_{1,5}$) for the right knee flexion

Method	$a^R_{1,1}$	$a_{1,2}$	$a_{1,3}$	$a_{1,4}$	$a^L_{1,5}$
A. Equidistant	-0.86	16.06	32.99	49.91	66.83
B. Equifrequent	-0.86	10.41	16.84	34.37	66.83
C. Mutual information	8.44	16.83	16.84	28.09	66.83
D. Q without $I (\beta = 0)$	8.44	13.61	22.70	33.35	59.86
E. Q with I $(\beta = 1)$	0.00	10.00	20.00	30.00	60.00

Table 2: MBF parameter vector $\boldsymbol{a}_1$ for different criterions

6 Conclusions

In this paper, the data-based automatic design of membership functions for fuzzy systems is described. The algorithm bases on information-theoretical measures which have to be modified for fuzzy systems. In this modification, the easy approach of summing up membership values to estimate possibility distributions fails even for a very simple example. A more sophisticated approach with a constrained optimization provides better results. In addition, further aims as a good interpretability of membership functions have been included into the design approach. As an example, a design problem for a medical expert system to classify gait events in the instrumented gait analysis has been discussed.

References

[1] Krone, A.; Kiendl, H.: Rule-Based Decision Analysis with Fuzzy-Rosa Method. In: *Europ. Workshop Fuzzy Decision Analysis for Management, Planning and Optimization EFDAN* (Felix, R., Hg.), S. 109–114. Dortmund: Fuzzy Demonstrations-Zentrum. 1996.

[2] Hayashi, I.; Maeda, T.; Bastian, A.; Jain, L. C.: Generation of Fuzzy Decision Trees by Fuzzy ID3 with Adjusting Mechanism of AND/OR Operators. In: *Proc. IEEE Int. Conf. Fuzzy Systems*, S. 681–685. Piscataway, NJ. 1998.

[3] Mikut, R.; Jäkel, J.; Malberg, H.; Bretthauer, G.: Datenbasierter Entwurf von Fuzzy-Systemen für medizinische Diagnoseaufgaben. *Automatisierungstechnik* 48 (2000) 7, S. 317–326.

[4] Jäkel, J.; Gröll, L.; Mikut, R.: Automatic Generation and Evaluation of Interpretable Fuzzy Rules. In: *New Frontiers in Computational Intelligence and its Applications* (Mohammadian, M., Hg.), S. 1–10. Amsterdam: IOS Press. 2000.

[5] Babuška, R.: *Fuzzy Modeling for Control*. Boston: Kluwer. 1998.

[6] Koch, M.; Kuhn, T.; Wernstedt, J.: *Fuzzy Control: optimale Nachbildung und Entwurf optimaler Entscheidungen*. München: Oldenbourg. 1996.

[7] Jäkel, J.: *Linguistische Fuzzy-Systeme mit verallgemeinerten Konklusionen und ihre Anwendung zur Modellbildung und Regelung*. Fortschritt-Berichte VDI, Reihe 8, Nr. 793. Düsseldorf: VDI-Verlag. 1999.

[8] Cios, K. J.; Pedrycz, W.; Swiniarski, R. W.: *Data Mining Methods for Knowledge Discovery*. Boston: Kluwer Academic. 1998.

[9] Perry, J.: *Gait Analysis. Normal and Pathological Function*. Thorofare: Slack Inc. 1992.

[10] Schablowski, M.; Rupp, R.; Gerner, H. J.: A Novel Approach to Gait Analysis on the Treadmill: II) Additional Requirements for Application in Case of Paraplegia. *Gait and Posture* 10 (1999), S. 83.

[11] Metaxiotis, D.; Siebel, A.; Döderlein, L.: Hip Deformities in Walking Patients with Cerebral Palsy. *Gait and Posture* 8 (1998), S. 64.

[12] Zlatnik, D.: Intelligently Controlled Above Knee (A/K) Prosthesis. In: *Proc. 4th Int. Conf. On Motion and Vibration Control (MOVIC'98)*. Zürich. 1998.

[13] Mikut, R.; Peter, N.; Bretthauer, G.; Rupp, R.; Abel, R.; Siebel, A.; Döderlein, A.; Gerner, H.-J.: Fuzzy-Regelgenerierung und Multivariate Statistische Verfahren zur Schritt-Phasenerkennung in der Instrumentellen Ganganalyse. In: *Berichtsband 11. Workshop* Fuzzy Control *d. GMA-UA 5.22*. (eingereichter Beitrag). 2000.

[14] Abel, R.; Rupp, R.; Siebel, A.; Döderlein, L.; Gerner, H. J.; Mikut, R.; Peter, N.; Bretthauer, G.: Classification of Gait Analysis Data Using Fuzzy Logic Based Rule Generation and Multivariate Statistical Analysis. In: *ESMAC'00*. Lund. 2000.

[15] Miller, G. A.: Note on the Bias of Information Estimates. In: *Information Theory in Psychology* (Quastler, H., Hg.). Glencoe, Illinois: Free Press. 1955.

[16] Rives, J.: FID3: Fuzzy Induction Decision Tree. In: *Proc. 1st Int. Symp. Uncertainty, Modelling and Analysis*, S. 457–462. Los Alamitos, Calif.: IEEE Computer Soc. Press. 1990.

[17] Fletcher, R.: *Practical Methods of Optimization*. Chichester: John Wiley & Sons. 1987.