Creation and Application of Fuzzy Rules to Access the Creditworthiness of Enterprises

Stefan Blochwitz and Judith Eigermann
Deutsche Bundesbank, Abt. Geldmarkt und Instrumentarium
Wilhelm–Epstein–Straße 14, D–60431 Frankfurt am Main Germany
email: {stefan.blochwitz, judith.eigermann}@bundesbank.de

Abstract

As practiced by banks, assessing the creditworthiness of an enterprise is a process that consists of weighting up several facts against each other. Whereas the collection of the facts is often standardized, the evaluation of them is carried out depending mainly on experience and internal standards of the banks by a human being called credit–analyst. The final decision on the creditworthiness is then made by a more senior officer of the bank.

Therefore, the mentioned process is well suitable for modelling with a fuzzy–system for supporting the decision. But for a successful practical implementation the system has to meet several criteria for success:

(i) Minimized classification error, i.e. the system should classify as many creditworthy or not creditworthy enterprises as possible as creditworthy or not creditworthy, respectively.

(ii) Transparency of the decision–process, i.e. an external analyst should be able to understand the decision.

(iii) Correctness of the decision, i.e. the external analyst should have got the same decision for at least similar reasons.

In our paper we discuss these criteria and their influence on the fuzzy–model, especially on the membership–functions.

KEYWORDS: membership functions, non linear programming, semantic correctness of rules, boundary conditions

1 Assessing an enterprise

1.1 The process and its modelling

Assessing the creditworthness of an enterprise is a process yielding to a Yes–No–decision: It is to decide whether the risk of lending a certain amount of money to an enterprise is acceptable or not, because the enterprise is able to pay back that money plus the interest. In a bank this process has to be performed many times, e.g. at the Deutsche Bundesbank approximately 65,000 times per year. Any automatization or standardization of this process will increase the effectiveness and is therefore highly appreciated.
In practice, this process is performed in four steps:

(i) Collecting the data used for assessment.

(ii) Applying a tool for preselection. This stage of assessment is necessary for separating the unambiguous cases from the ambiguous ones. Whereas the former enterprises can be regarded as assessed sufficiently, the latter ones need a further assessment. A great deal of work is done in this subject, see [1]–[6].

Mostly, in this stage of assessment the enterprises are classified by a score called numeric value Z and it holds then:

No. of case	Value of Z	Meaning for assessment
1	$Z \leq Z_{\text{bad}}$	enterprise is probably not creditworthy
2	$Z \geq Z_{\text{good}}$	enterprise is probably creditworthy
3	$Z_{\text{bad}} < Z < Z_{\text{good}}$	enterprise may be either creditworthy or not

Cases 1 and 2 are the unambiguous ones, 3 is the ambiguous one.

(iii) Revision of the results of step (ii). Whereas for reasons inherit to the method applied in the step before the classification of enterprises with score under Z_{bad} or above Z_{good} is highly sure, the rest of the enterprises needs further examination. Therefore the process focuses now on these enterprises.

In this stage the credit–analyst is challenged. He — or she — goes on and looks for facts, that are not yet regarded. The analyst bears in mind a few rules and checks with their help the state of the enterprise.

Keeping in mind that the score is intended as measure for the creditstanding then all these rules can be formulated in a form like:

> **If** property X_1 is observed **and if** property X_2 is observed **and if** $\cdots$ **and if** property X_{N} is observed **then** the enterprise's score has to be increased (or decreased).

An example of such a rule is

> **If** the turnover is much lower than in the previous year **then** the enterprise's score has to be decreased.

Often this step has to be performed by the credit analyst, although it would be wishful to standardize this step too. Recent research goes into this direction[9], [10].

(iv) Final assessment basing on the results of steps (ii) and (iii). In practice it is observed that for the same enterprise some rules are relevant which plea for decreasing and some that plea for increasing of the score. It is necessary to weigh up these contradicting rules in order to get a unambiguos decision. The credit–analyst comes to this final decision by its experience.

In our opinion the decision–process of step (iii) can be modelled by a fuzzy–system. The i–th rule writes then:

> **If** X_{i1} is A_{i1} **and** $\cdots$ **and** $X_{i\text{N}}$ is $A_{i\text{N}}$ **then** change Z by ΔZ.

Here $X_{i1} \cdots X_{i\text{N}}$ are the input variables and $A_{i1} \cdots A_{i\text{N}}$ are fuzzy–sets defined by a fuzzy partition of the definition range of these input variables. Our goal is then to assign the creditworthy enterprises a score $Z > Z_{\text{good}}$ and the not creditworthy enterprises a score $Z < Z_{\text{bad}}$.
Therefore it is suitable to calculate the aggregated output of the whole system as

$$Z^{\text{new}} = Z + \frac{\Delta Z^{+} \cdot \sum\limits_{\text{pro–rules}} \gamma_i \prod\limits_{i\text{–th rule}} \mu_{A_{ij}}(X_{ij}; p_{ij,1}, \cdots p_{ij,\text{N}}) + \Delta Z^{-} \cdot \sum\limits_{\text{contra–rules}} \gamma_i \prod\limits_{i\text{–th rule}} \mu_{A_{ij}}(X_{ij}; p_{ij,1}, \cdots p_{ij,\text{N}})}{\sum\limits_{\text{all rules}} \gamma_i \prod\limits_{i\text{–th rule}} \mu_{A_{ij}}(X_{ij}; p_{ij,1}, \cdots p_{ij,\text{N}})}, \tag{1}$$

with ΔZ^{+} as the maximum surcharge to the Z–score, ΔZ^{-} as the the maximum deduction of the Z–score, γ_i as the global weight of the i–th rule, and $\prod\limits_{i\text{–th rule}} \mu_{A_{ij}}(X_{ij}; p_{ij,1}, \cdots p_{ij,\text{N}})$ as the degree of activation of the i–th rule.
This degree of activation depends only on the membership–functions $\mu_{A_{ij}}$ of the fuzzy partition of the relevant input variables X_{ij}. The parameters $p_{ij,1}, \cdots, p_{ij,\text{N}}$ stand for positional parameters of the membership–functions and are discussed in detail in Sec. 2.2.

1.2 Criteria for success

The first and most essential criterion for success are the apparent successful classification rate (ASCR) and the rate of ambiguous classified enterprises (ACER) of the system. These quantities are defined in analogy to statistical measures of performance[11]. Both quantities can be estimated from a representative sample drawn from the whole population of enterprises. Important is, that it is known for the members of this sample in advance, whether they are creditworthy or not[1].

[1] So an enterprise may be regarded as not creditworthy if an application has been made for the instigation of bankruptcy or composition proceedings against it, otherwise it may be regarded as creditworthy.

Let the for the creditworthiness of i–th sample–member $T_i^{\text{obs.}}$ hold $T_i^{\text{obs.}} = \begin{cases} Z_{\text{good}} & \text{if it is creditworthy} \\ Z_{\text{bad}} & \text{if it is not creditworthy} \end{cases}$ then is valid for ASCR [2] and ACER with $\Theta(x)$ as Heaviside's function[3], $P_{\text{ap}}^{\text{cw}}$ and $P_{\text{ap}}^{\text{ncw}}$ as the prior probabilities for creditworthy and not creditworthy enterprises, respectively:

$$\text{ASCR} = \frac{\sum\limits_{\text{sample}} \left[P_{\text{ap}}^{\text{cw}} \overbrace{\Theta(Z_i^{\text{new}} - Z_{\text{good}})\Theta(T_i^{\text{obs.}} - Z_{\text{good}})}^{\text{creditworthy, classified as creditworthy}} + P_{\text{ap}}^{\text{ncw}} \overbrace{\Theta(Z_{\text{bad}} - Z_i^{\text{new}})\Theta(Z_{\text{bad}} - T_i^{\text{obs.}})}^{\text{not creditworthy, classified as not creditworthy}} \right]}{\sum\limits_{\text{sample}} \left[P_{\text{ap}}^{\text{cw}} \Theta(T_i^{\text{obs.}} - Z_{\text{good}}) + P_{\text{ap}}^{\text{ncw}} \Theta(Z_{\text{bad}} - T_i^{\text{obs.}}) \right]} \tag{2}$$

$$\text{ACER} = \frac{\sum\limits_{\text{sample}} \left[\Theta(Z_i^{\text{new}} - Z_{\text{bad}})\Theta(Z_{\text{good}} - Z_i^{\text{new}}) \cdot \left(P_{\text{ap}}^{\text{cw}} \Theta(T_i^{\text{obs.}} - Z_{\text{good}}) + P_{\text{ap}}^{\text{ncw}} \Theta(Z_{\text{bad}} - T_i^{\text{obs.}}) \right) \right]}{\sum\limits_{\text{sample}} \left[P_{\text{ap}}^{\text{cw}} \Theta(T_i^{\text{obs.}} - Z_{\text{good}}) + P_{\text{ap}}^{\text{ncw}} \Theta(Z_{\text{bad}} - T_i^{\text{obs.}}) \right]} \tag{3}$$

Equation (1) allows it to determine the amount of which a specific rule contributes to the new Z–score. The contribution C_i of the i–th rule is

$$C_i = \frac{\gamma_i \cdot \prod\limits_{i\text{–th rule}} \mu_{A_{ij}}(X_{ij}; p_{ij,1}, \cdots p_{ij,\text{N}})}{\sum\limits_{\text{all rules}} \gamma_i \prod\limits_{i\text{–th rule}} \mu_{A_{ij}}(X_{ij}; p_{ij,1}, \cdots p_{ij,\text{N}})} \cdot \begin{cases} \Delta Z^{+} & \text{if the rule is 'pro'} \\ \Delta Z^{-} & \text{if the rule is 'contra'} \end{cases} \tag{4}$$

The second and third criterion for success deal with these contributions.
The next criterion is transparency of the decision. C_i makes the decision–process transparent to an external analyst, because the bigger the absolute amount $|C_i|$ is the more important is the i–th rule for Z^{new}. So the second criterion is already fullfilled by Eq. (1) as the model of the analyst's decision process.
The last but not least criterion is the correctness of the decision. That means that out of all pro– and contra–rules only rules which would have been chosen by an experienced credit–analyst, too, contribute to Z^{new}. This criterion has implications on the fuzzy–partition, what is discussed in detail in Sec. 2.2.

2 Adjusting the model to reality

2.1 Fuzzy–system and reality

The fuzzy–system should reflect reality. Basically, there are two opportunities for implementing reality:

(i) The pro– and contra–rules that are included in the model (1).

(ii) The way the definition-range of the inputvariables X_{ij} is fuzzy–partitioned.

Our assumption is — as mentioned in Sec. 1.2 —, that a representative sample exists. In this sample it can be tested, whether a certain rule is relevant for the sample or not.
There are several methods, how to estimate the relevance of a rule on the basis of a sample. One approach is according to the Fuzzy–ROSA method [7], which subjects every rule to a test, founded on statistical evaluation and confidence intervals. Each data set of the sample contributes in eq. (1) to the final result of the evaluation according to the degrees of matching of its premises. Consider with N as the number of data sets in the sample, $\Pr(i\text{–th rule}) = \frac{\sum\limits_{k=1}^{\text{N}} \prod\limits_{i\text{–th rule}} \mu_{A_{ij}}(X_{ij}; p_{ij,1}, \cdots p_{ij,\text{N}})}{\text{N}}$ as the unconstrained probability of

[2] Note, that enterprises classified into the ambiguos region do not count for ASCR.

[3] This function is defined as $\Theta(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$.

the i–th rule, $\Pr(i\text{–th rule} \mid \text{Conclusion Th}_i) = \dfrac{\sum_{k=1}^{N} \prod_{i\text{–th rule}} \mu_{A_{ij}}(X_{ij}; p_{ij,1}, \cdots p_{ij,N}) \cdot \Theta(-|T_k^{\text{obs.}} - \text{Th}_i|)}{\sum_{k=1}^{N} \prod_{i\text{–th rule}} \mu_{A_{ij}}(X_{ij}; p_{ij,1}, \cdots p_{ij,N})}$ as the constrained probability of the conclusion part[4] of i–th rule. Based on these values a rule is relevant if the constrained probability exceeds the unconstrained probability for finding a good or bad enterprise in the given sample: $\Pr(i\text{–th rule} \mid \text{Conclusion Th}_i) > P_{\text{ap}}^{\text{cw}}\Theta(\text{Th}_i - Z_{\text{good}}) + P_{\text{ap}}^{\text{ncw}}\Theta(Z_{\text{bad}} - \text{Th}_i)$. The significance of estimation is also considered by calculating confidence intervals, which must not overlap[8]. A more rough approach is to accept only rules for which holds $\Pr(i\text{–th rule}) > P_{\text{sig}}$[5] and $\Pr(i\text{–th rule} \mid \text{Conclusion Th}_i) > 0.5$.
So it can be tested, starting from all rules that are proposed by human experience, which of these rules can in fact be applied. But it can be done the other way round, too: Starting from a universe of all possible rules[6] all rules, that are relevant for the sample may be gotten. In a second step these relevant rules may be filtered by human expert(s).

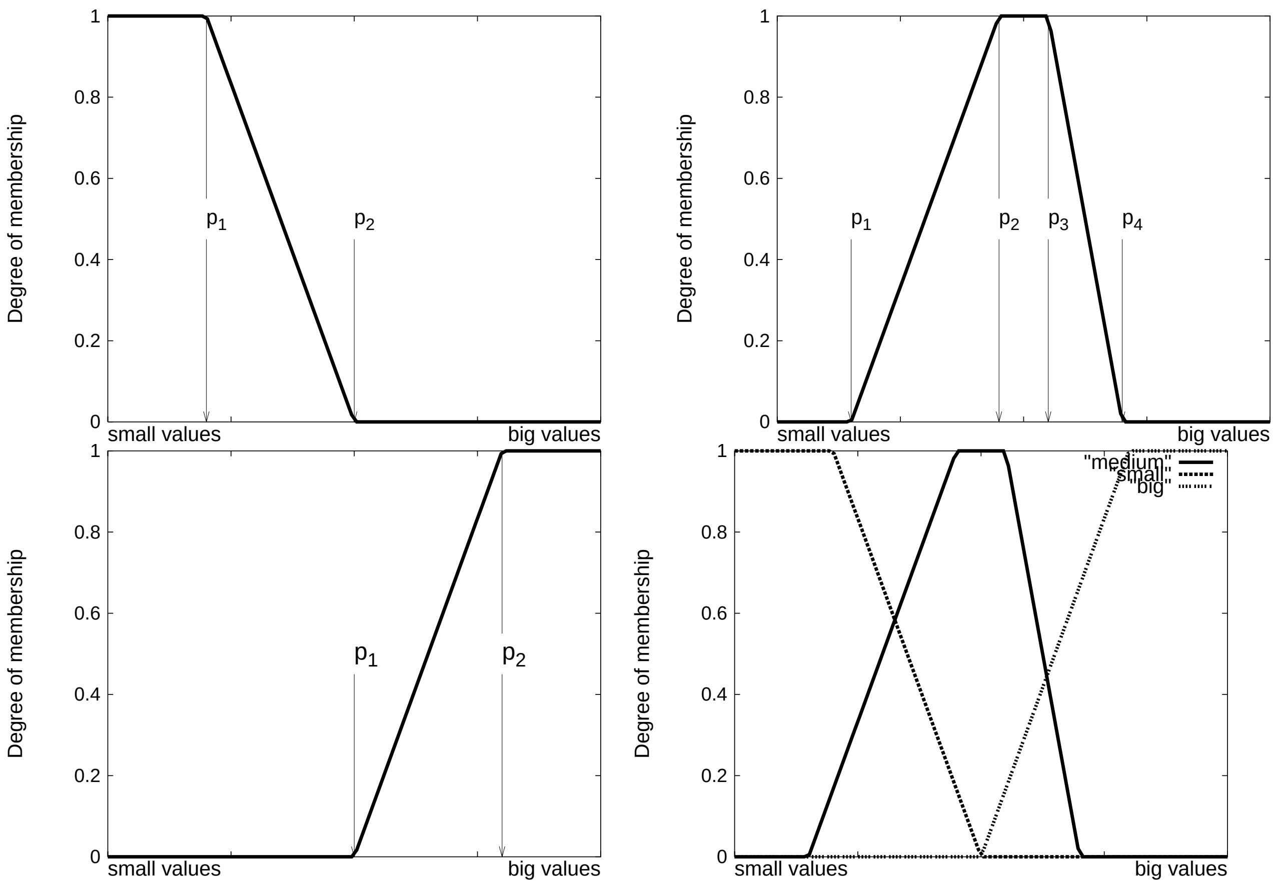

Figure 1: The 3 basic–types of membership–functions for the linguistic terms 'small', (upper left) 'medium' (upper right) and 'big' (lower left) and an example for a fuzzy–partition (lower right). Note, the parameters $p_1, \cdots, p_4$ allow to control the position of the membership–functions.

Anyway, both methods result in a set of relevant and accepted rules. The second method may be yield more rules, because it guarantees, that no rule is forgotten. Both methods require a fairly good fuzzy–partition of the input space.
Every linguistic variable can be modelled by a combination of the membership–functions plotted in Fig. 1. In our opinion these piecewise linear functions are good enough; more sophisticated functions e.g. Gauss'

[4]For the conclusion Th_i of the i–th rule it is supposed that holds $\text{Th}_i = \begin{cases} Z_{\text{good}} & \text{if the } i\text{–th rule increases } Z \\ Z_{\text{bad}} & \text{if the } i\text{–th rule decreases } Z \end{cases}$.

[5]This condition secures a sufficient amount of data, for which the regarded rule is relevant.

[6]This universe results from all possible combinations of the fuzzy–input–space.

function for the term 'medium' do not really a better work. For a rough estimation of the performance of the fuzzy–system (1) these membership–funtions are well suitable.
All the membership–functions have a different set of parameters $p_1, \cdots$ that allows to control their actual position in the input space. For an optimal performance of system (1) it is important to have an optimal position — and thus an optimal set of positional parameters — of every membership–function. This optimal set of parameters can be determined either by human experience or by adopting the parameters in a more objective way. A method that is performed succesfully at the Deutsche Bundesbank is introduced in Sec. 2.3.

2.2 Semantic correctness of the rules

Let us consider the following three rules:

(i) **If** turnover is much lower than in the previous year **then** the enterprise's score has to be decreased.

(ii) **If** turnover is unchanged or moderately higher than in the previous year **then** the enterprise's score has to be decreased.

(iii) **If** turnover is much higher than in the previous year **then** the enterprise's score has to be decreased.

The inputvariable 'turnover' has to be partitioned into three features: *much lower*, *unchanged or moderately higher* and *much higher*. By this partition it has to be obeyed that the partition is reasonable.
According to fuzzy set theory an overlapping of the membership–function is possible and an important feature of these functions. But not every partition is sensible: An overlapping of the membership–functions for the terms *much lower* and *much higher* contradicts human thinking and it is therefore not explainable (cmp. also Fig. 2).
Remember, that we in Eq. (4) the contributions C_i introduced as quantity to make the decision of the system transparent. Overlapping membership–functions result in rule–contributions for both rules in the area of overlap and thus the result is explained in a not understandable way. Although this overlapping may improve ASCR or ACER, it should be avoided, since transparency and explainable decisions of the system are as important as a good ASCR and ACER.
A fuzzy–partition of the input space that allows only acceptable areas of overlappings is called semantically correct.

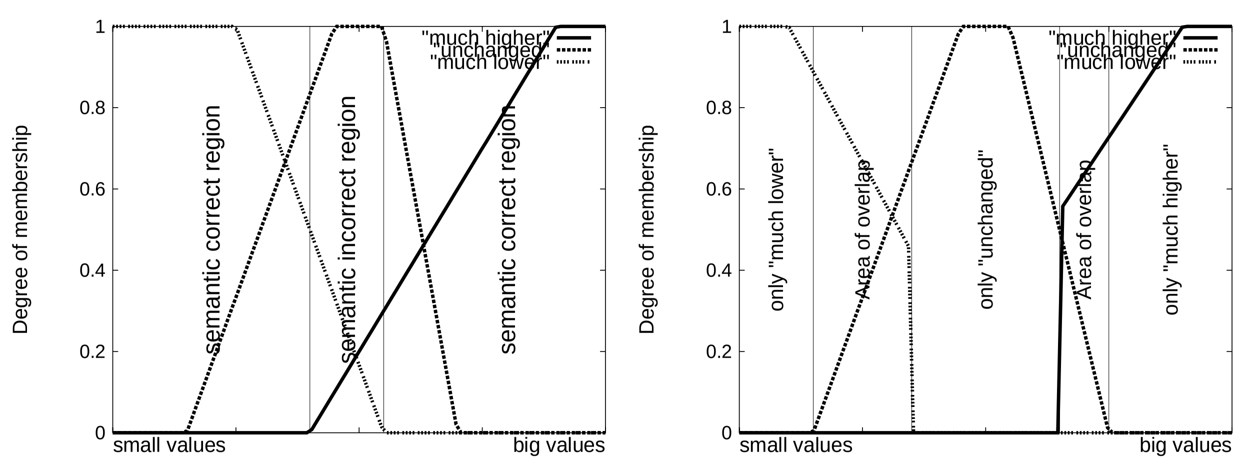

Figure 2: An example for semantic incorrect fuzzy–partition (left) of the input space with the membership–functions from Fig. 1. Semantic correct partition of input–space by setting membership–functions to 0 (right).

As a result of these considerations we end up (for an inputvariable that is splitted into three linguistic terms) for the semantically correct partition with five different subsets (cmp. Fig. 2):

- three subsets of clear–cut membership in which the inputvariable — with a varying degree of membership — belongs exclusively to one linguistic term (in our example: *much lower*, *unchanged or moderately higher*, or *much higher*), and

- two areas of overlap in which a fluid transition occurs between the clear–cut subsets because in them the feature belongs to two linguistic terms (in our example: *much lower* and *unchanged or moderately higher* or *unchanged or moderately higher* and *much higher*).

Semantic correctness can easily introduced by an additional parameter per membership–function in conjunction with boundary conditions. The new parameter is called $p_{\min}$ for the membership–function for *much higher* and is called $p_{\max}$ for *much lower*. It gives the maximum or minimum value of the inputvariable, respectively, for which the membership–function is greater than zero. As an example the membership–function for *much higher* is given by:

$$\mu_{\text{much higher}}(X, p_1, p_2, p_{\min}) = \begin{cases} 0 & \text{for } X < \max(p_1, p_{\min}) \\ 1 & \text{for } X > \max(p_2, p_{\min}) \\ \frac{X - p_1}{p_2 - p_1} & \text{elsewhere} \end{cases}$$

The concept for the much–higher–membership–function is adoptable to all kinds of membership–functions. And since it is easier to formulate sensible ranges for the parameters for a human expert semantic correctness can be guaranteed by boundary conditions.
Boundary conditions are the mathematical translation for semantic correctness. To give an example: For the membership–functions for *unchanged or moderately higher* and *much higher* it is demanded from the point of view of a human expert that holds:

(i) The degree of membership of input–values $X < 0$ to *much higher* has to be zero.

(ii) The degree of membership of input–values to *much higher* where the degree of membership to *unchanged or moderately higher* is greater than .5 has to be zero, too.

Thus the resulting boundary conditions for the parameters of the function for *much higher* $p_1^{\text{much higher}}$, $p_2^{\text{much higher}}$ and $p_{\min}^{\text{much higher}}$ and for the parameters of the function for *unchanged or moderately higher* $p_3^{\text{unchanged}}$ and $p_4^{\text{unchanged}}$ are $p_1^{\text{much higher}} > 0$, $p_{\min}^{\text{much higher}} > 0$ and $p_{\min}^{\text{much higher}} \geq .5(3p_3^{\text{unchanged}} - p_4^{\text{unchanged}})$.
According to our experience all these necessary boundary conditions plus external knowledge about the membership–function[7] can be transformed into a linear system. Thus, a linear system of inequality constraints like eqns. (5) guarantee semantic correctness:

$$\begin{array}{lll} a_1 & \{\leq=\} & \sum_{\text{All parameters}} a_{ijk1} p_{ij,k} \\ & \vdots & \\ a_N & \{\leq=\} & \sum_{\text{All parameters}} a_{ijkN} p_{ij,k} \end{array} \tag{5}$$

2.3 Adjusting as a NLP–Problem

With the knowledge of Secs. 2.1 and 2.2 we can rewrite our goal to adjust the fuzzy–system (1) to reality as:

> For a representative sample of creditworthy and not creditworthy enterprises maximize the function (2) and minimize (3) with Z given by eq. (1) under the linear constraints (5) with respect to the variables $p_{ij,k}$, γ_i, ΔZ^+ and ΔZ^-.

This is a nonlinear–programming–problem and can be solved by numerical techniques developed for these problems[12], [13]. After having the solution — if it exists — we can say, that our fuzzy–system is adjusted in the most possible way to reality. The amount of free parameters determines the size of the sample in order to avoid overfitting: Ten Inputvariables and three fuzzysets per variable result in 80 parameters, each rule in a further parameter. So we may end up in a system with something between 150 and 200 parameters.

[7] This knowlwdge reflects the experience of the human analyst. So it may be generally accepted that a *much–higher*–turnover starts at 10%, what means in terms of parameters of the membership–function that $p_{\min}^{\text{much higher}} > 0.10$ has to be valid.

Another consideration deals with the functions we wish to minimize. Instead of minimizing two functions it is much better to minimize only one. Keeping in mind that it is worse to classify an enterprise false than to classify it not we can in eq. (2) approximate Heaviside's function with an argument containing Z_i^{new} by a Fermi– or logistic function with the same argument. For creditworthy enterprises that function is 1 if $Z^{\mathrm{new}} > Z_{\mathrm{good}}$, 0 if $Z^{\mathrm{new}} < Z_{\mathrm{bad}}$ and between 0 and 1 if $Z_{\mathrm{bad}} \leq Z^{\mathrm{new}} \leq Z_{\mathrm{good}}$ and vice versa for the not creditworthy enterprises — cmp. Fig. 3.

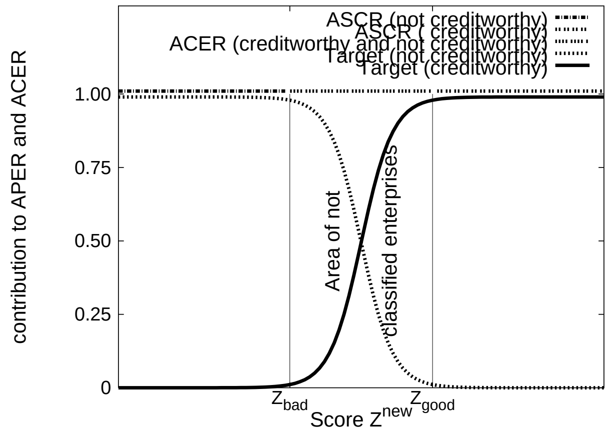

Figure 3: ASCR, ACER and Targetfunctions for creditworthy and not creditworthy enterprises.

We then can forget about ACER, since to the 'new' ASCR contributes every correct classified enterprise with 1, every false classified enterprise with 0 and every not classified enterprise with a value between 0 and 1. That value is larger Z^{new} is the closer Z^{new} is to the boundaries for correct classification. Therefore the new target is maximized if as many enterprises as possible are classified correctly. The NLP–Problem reduces then to maximize only one function.

3 Results

The principles explained in Sec. 2 are implemented successfully for more than a year in the Bundesbank's method of assessing the creditworthiness of business enterprises[14]. The score Z derived from the statistical tool for preselection comes from a discriminant analysis. The here presented methodolgy of rule generation and rule testing yields rules that are well acceppted by the credit analyst.

	ASCR in %	ACER in %
after step (ii), i.e. Discriminant analysis	70.9	17.5
after Fuzzy–model of step (iii)	78.4	5.8

Table 1: Classification results in the standardized process of assessing the creditworthiness at the Deutsche Bundesbank. Note, that these results for reasons of comparability are determined out of the sample under the assumption of equal prior probabilities ($P_{\mathrm{ap}}^{\mathrm{ncw}} = P_{\mathrm{ap}}^{\mathrm{cw}} = .5$).

Table 1 shows, that the main goal is reached: ASCR has increased, whereas ACER decreased.

References

[1] Baetge, J.: Möglichkeiten der Früherkennung negativer Unternehmensentwicklungen mit Hilfe statistischer Jahresabschlußanalysen, in: Zfbf (1989), 41. Jg., pp. 792-810

[2] Baetge, J.; Beuter, H. B.; Feidecker, M.: Kreditwürdigkeitsprüfung mit Diskriminanzanalyse, in: Die Wirtschaftsprüfung, 45. Jg (1995)., pp. 749-761

[3] Fahrmeir, L.; Frank, M.; Hornsteiner, U.: Bonitätsanalyse mit alternativen Methoden der Diskriminanzanalyse, in: Die Bank (1996), pp. 368-373

[4] Kühnenberger, M.; Eckstein, P.; Woithe, M.: Die Diskriminanzanalyse als ein Instrument zur Früherkennung negativer Unternehmensentwicklungen, in: Zeitschrift für Betriebswirtschaft, 66. Jg. (1996), S.1449-1464

[5] Blochwitz, S.; Eigermann, J.: Effiziente Kreditrisikobeurteilung durch Diskriminanzanalyse mit qualitativen Merkmalen to appear in: Handbuch Kreditrisikomodelle und –derivate, hrsg. von Roland Eller und Walter Gruber, 1999

[6] Blochwitz, S.; Eigermann, J.: Unternehmensbeurteilung durch Diskriminanzanalyse mit qualitativen Merkmalen to appear in zfbf (1st quarter 2000)

[7] Kiendl, H.: Fuzzy Control methodenorientiert, R. Oldenbourg Verlag München Wien 1997

[8] Krone, A.; Schwane, U.: Generating Fuzzy Rules from Contradictory Data of Different Control Strategies and Control Performances in:Proc. 5th IEEE International Conference on Fuzzy Systems, New Orleans 1996, pp. 492–497

[9] Zimmermann, H.–J.: Fuzzy Set Theorie in: Wirtschaftswissenschaftliches Studium 28. Jg. (1999), pp. 22–29

[10] Wolz, M.: Die Krisenwarnfunktion des Abschlußprüfers, Gabler 1996

[11] Johnson, R. A.; Wichern, D. W.: Applied Multivariate Statistical Analysis, Prentice Hall, 3rd Edition 1992

[12] Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P.: Numerical Recipies in C, Cambridge University Press, 2nd edition 1995

[13] SAS Institute Inc.: SAS/OR Technical Report: The NLP–Procedure, Cary, NC (1997)

[14] The Bundesbank's method of assessing the creditworthiness of business enterprises in: Deutsche Bundesbank Monthly Report, January 1999 pp. 51–63